Announcements

Assignments

- HW8 - Due 3/28, 10 pm
- P4: MDP/RL - Due Thurs 11/17, 10 pm

Midterm 2

- One week!!! in lecture
- See Piazza post for details
- Review session 6-8pm Tuesday 3/28 in Rashid Auditorium
- In scope: Bayes nets representation and independence (today)

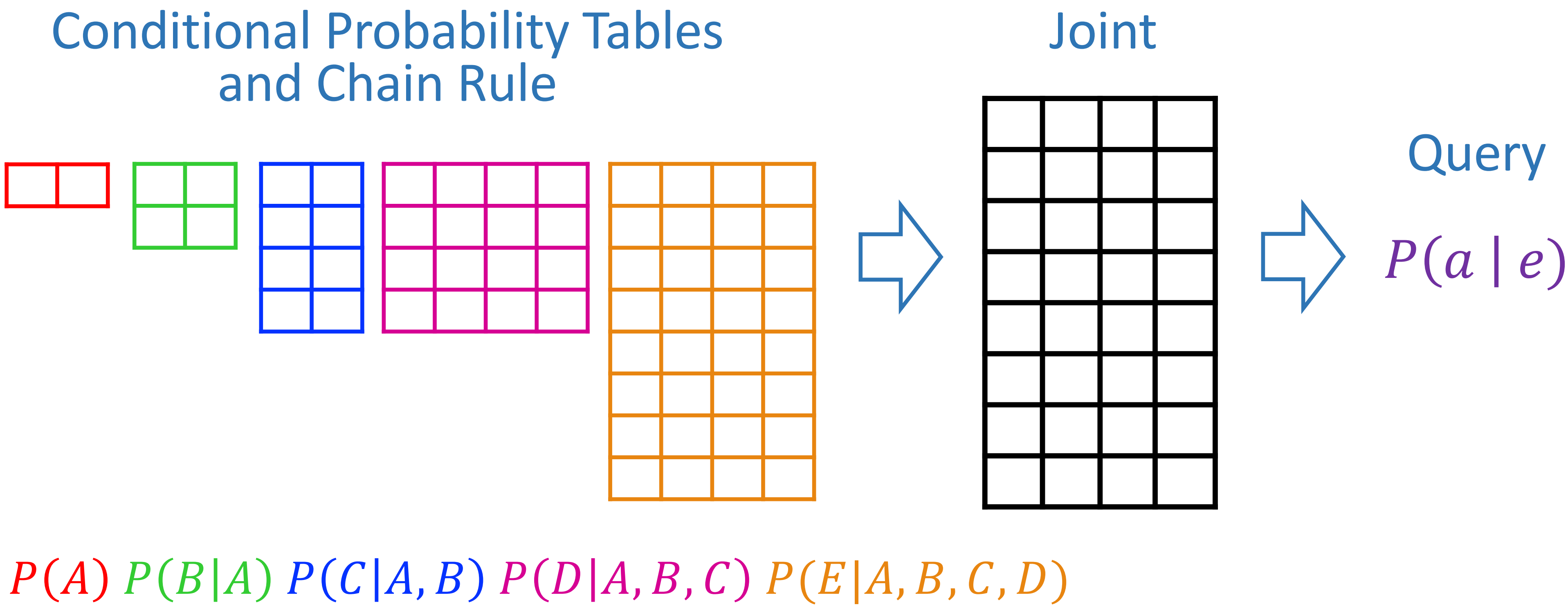
AI: Representation and Problem Solving

Bayes Nets: Independence

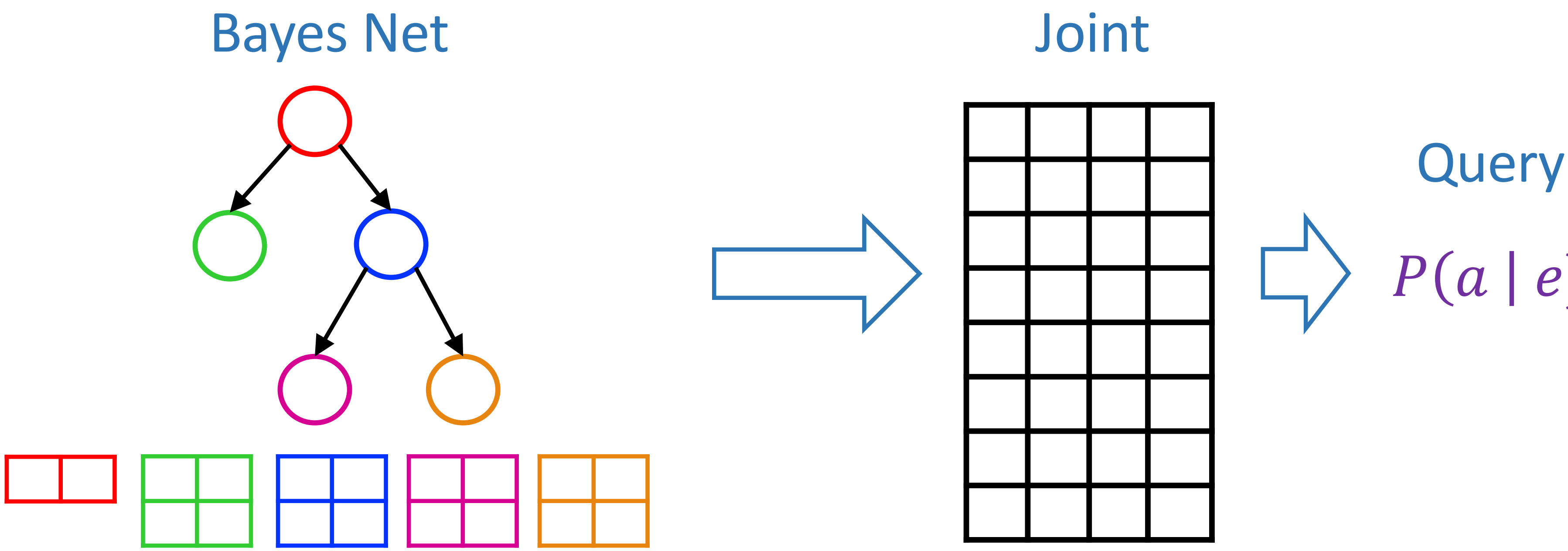
Instructor: Stephanie Rosenthal

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

Answer Any Query from Condition Probability Tables



Answer Any Query from Condition Probability Tables



$$P(a \mid e)$$

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

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

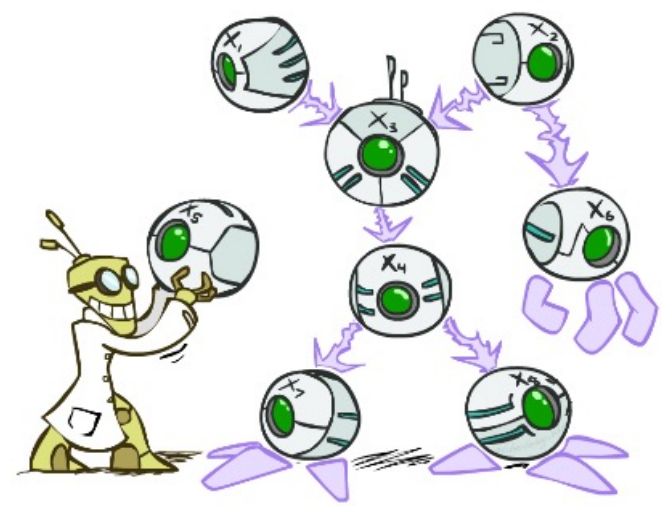
Bayes' Nets: Big Picture

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)

- A type of probabilistic graphical models
- We describe how variables locally interact
- Local interactions chain together to give global, indirect interactions

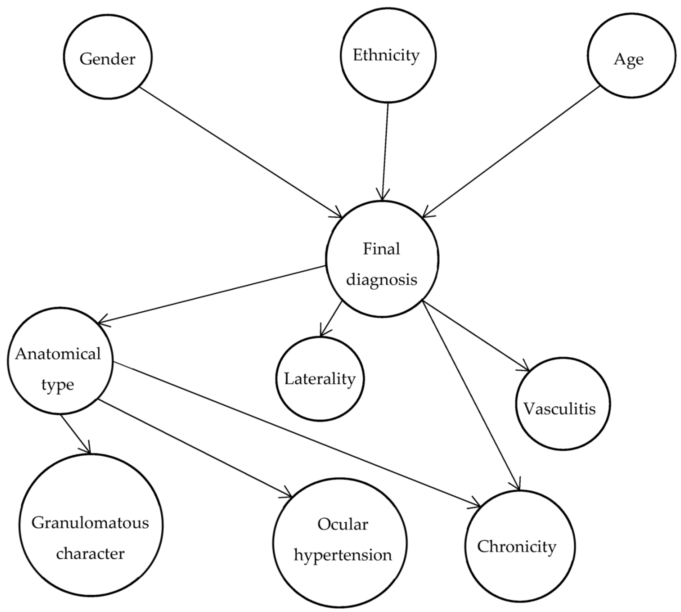


Bayes Nets for Medical Diagnosis

https://www.microsoft.com/en-us/research/people/dabelgra/

Developmental Profiles of Eczema, Wheeze, and Rhinitis:
Two Population-Based Birth Cohort Studies
Danielle Belgrave, et al. *PLOS Medicine*, 2014
https://journals.plos.org/plosmedicine/article?id=10.1371/journal.pmed.1001748

Bayes Nets for Medical Diagnosis



Jamilloux, Yvan, Nicolas Romain-Scelle, Muriel Rabilloud, Coralie Morel,
Laurent Kodjikian, Delphine Maucort-Boulch, Philip Bielefeld, and Pascal Sève. 2021.
"Development and Validation of a Bayesian Network for Supporting the Etiological Diagnosis of Uveitis"
Journal of Clinical Medicine 10, no. 15: 3398.

Bayes Nets for Medical Diagnosis

ELSEVIER

Biocybernetics and Biomedical Engineering

Volume 40, Issue 4, October–December 2020, Pages 1436-1445

Original Research Article

Computer-aided detection of COVID-19 from X-ray images using multi-CNN and Bayesnet classifier

Bejoy Abraham [a], Madhu S. Nair [b]

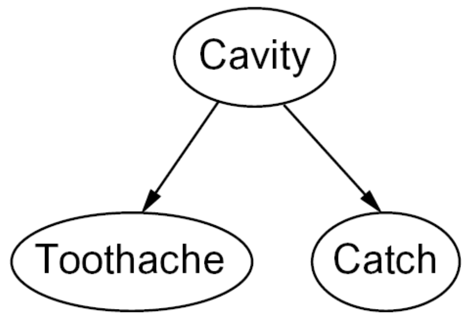
[a] Department of Computer Science and Engineering, College of Engineering Perumon, Kollam, Kerala, India

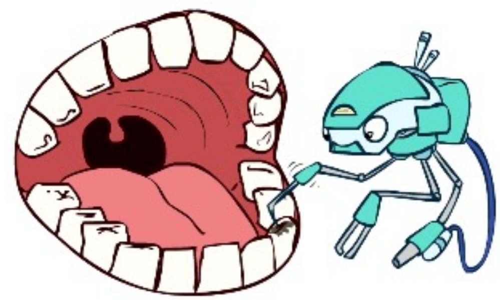
[b] Artificial Intelligence & Computer Vision Lab, Department of Computer Science, Cochin University of Science and Technology, Kochi, Kerala, India

Graphical Model Notation

Nodes: variables (with domains)

- Can be assigned (observed) or unassigned (unobserved)
- Observed does not mean $Variable = true$! Observed just means that we will have the value for that variable





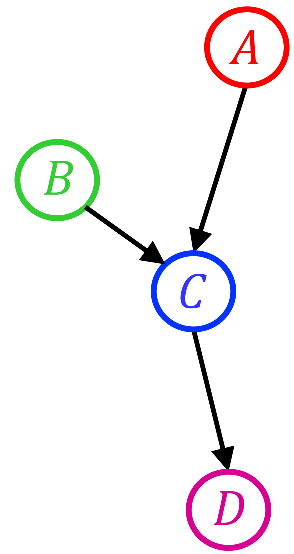
Edges

- Indicate "direct influence" between variables
- Absence of edges: encode conditional independence

For now: imagine that arrows mean direct causation (in general, they don't!)

Bayesian Networks

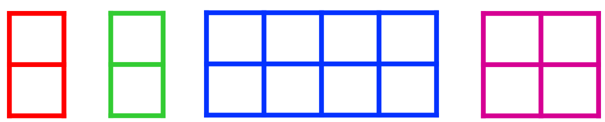
Bayes net



One node per random variable

Directed-Acyclic-Graph

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



Joint Probability

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

Encode joint distributions as product of conditional distributions on each variable

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

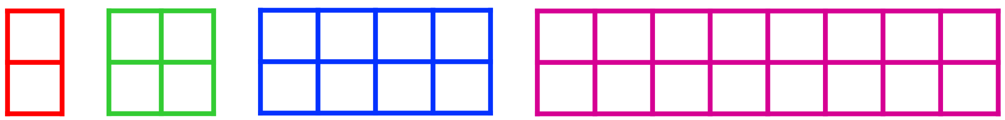
Bayesian Networks

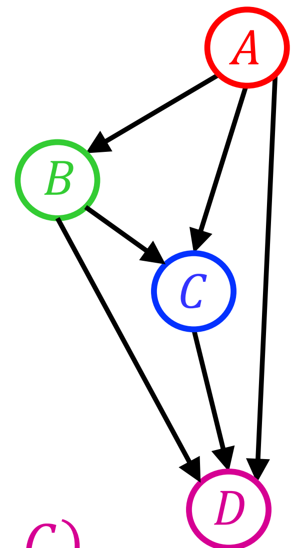
Bayes net

One node per random variable

Directed-Acyclic-Graph

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





Joint Probability

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

Encode joint distributions as product of conditional distributions on each variable

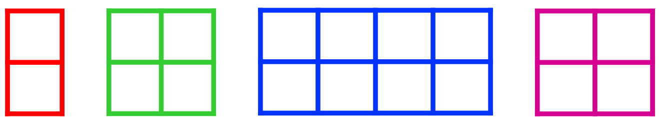
$$P(X_1, \dots, X_N) = \prod_i P(X_i |\ Parents(X_i))$$

Bayesian Networks

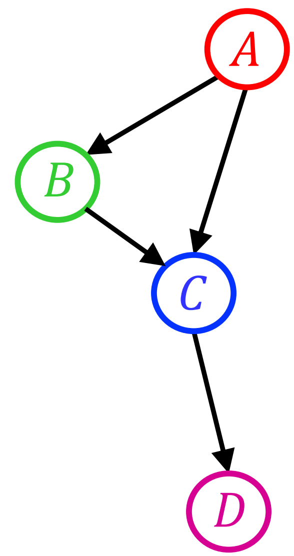
One node per random variable

Directed-Acyclic-Graph

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



Bayes net



Joint Probability

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

Encode joint distributions as product of conditional distributions on each variable

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

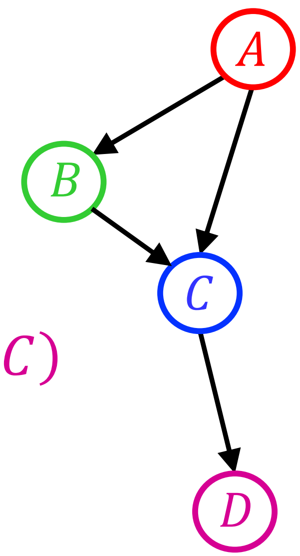
Bayesian Networks

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

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

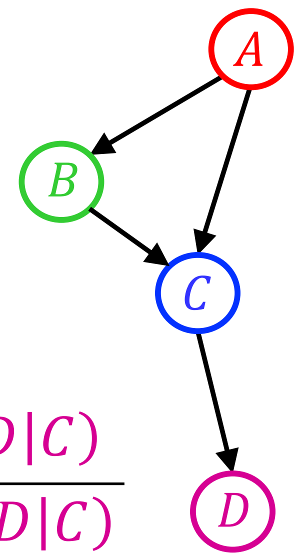
$$= \sum_{d \in D} P(A)\, P(B|A)\, P(C|A,B)\, P(D|C)$$

Bayes net



Bayesian Networks

Bayes net

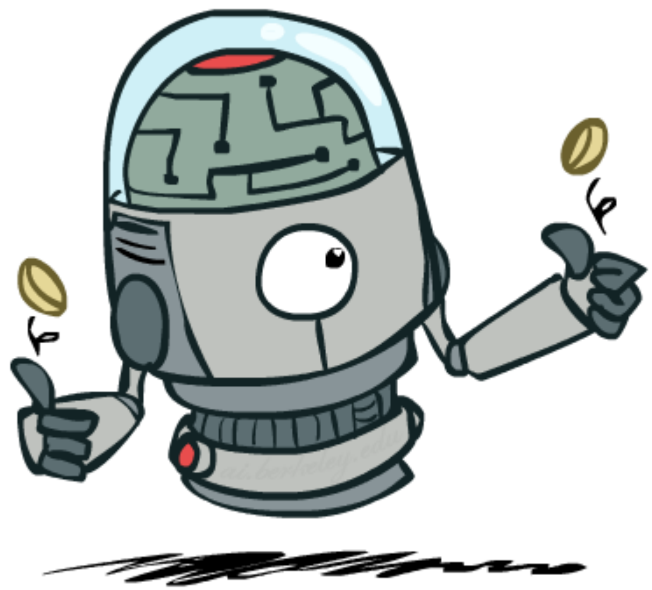


$$P(A,B,C,D) = P(A)\,P(B|A)\,P(C|A,B)\,P(D|C)$$

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

$$= \frac{\sum_{a \in A} P(A)\,P(B|A)\,P(C|A,B)\,P(D|C)}{\sum_A \sum_B P(A)\,P(B|A)\,P(C|A,B)\,P(D|C)}$$

Independence



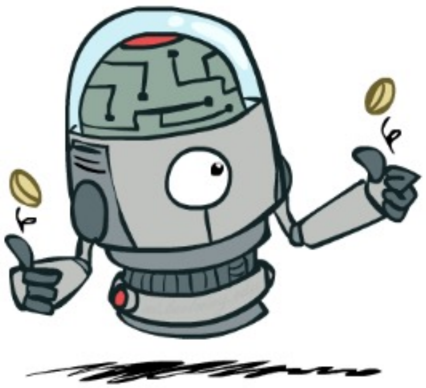
Independence

Two 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
- 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

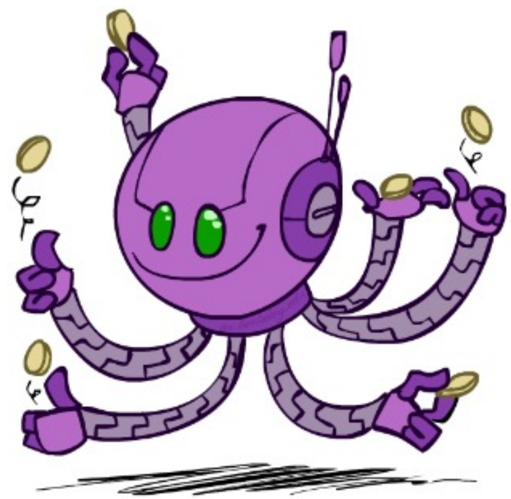
$P(X_2)$

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

Poll 1

Are T and W independent?

$P(T)$

T	P
hot	0.5
cold	0.5

$P(T,W)$

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

$P(W)$

W	P
sun	0.6
rain	0.4

Poll 1

Are T and W independent?

No

$P(T)$

T	P
hot	0.5
cold	0.5

$P(T,W)$

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

$P(T)P(W)$

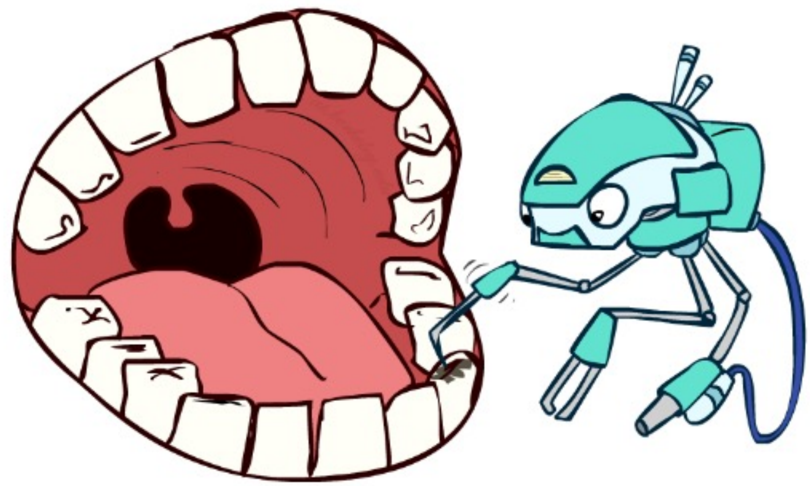
T	W	P
hot	sun	0.3
hot	rain	0.2
cold	sun	0.3
cold	rain	0.2

$P(W)$

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 | +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

Absolute (unconditional) 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

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

Independence Rules

- Independence

  If A and B are independent, then:

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

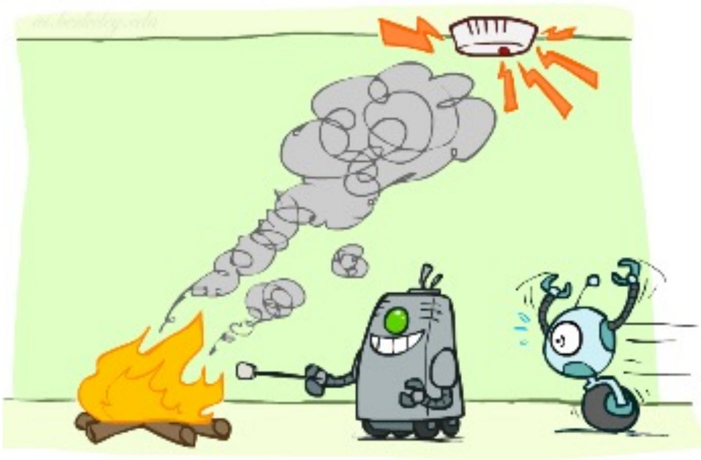
- Conditional independence

  If A and B are conditionally independent given C, then:

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

Conditional Independence and Bayes Nets

Fire, Smoke, Alarm



- Causal story to create Bayes net

- Assumptions

- Joint distribution

Conditional Independence and Bayes Nets

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:

$P(Rain, Traffic, Umbrella) =$

Bayes nets / graphical models help us express conditional independence assumptions

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:

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

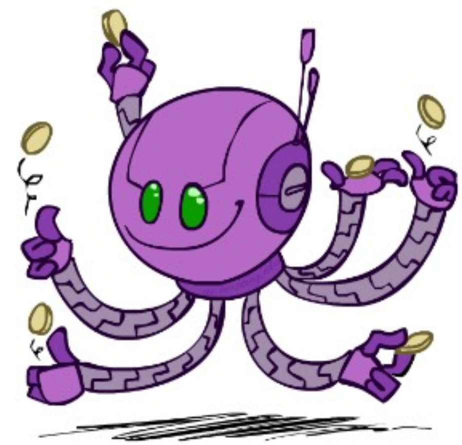
With assumption of conditional independence:

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

Bayes nets / graphical models help us express conditional independence assumptions

Example: Coin Flips

N independent coin flips



No interactions between variables: absolute independence

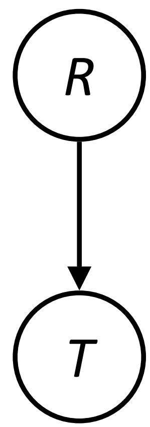
Example: Traffic

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 Bayes net!

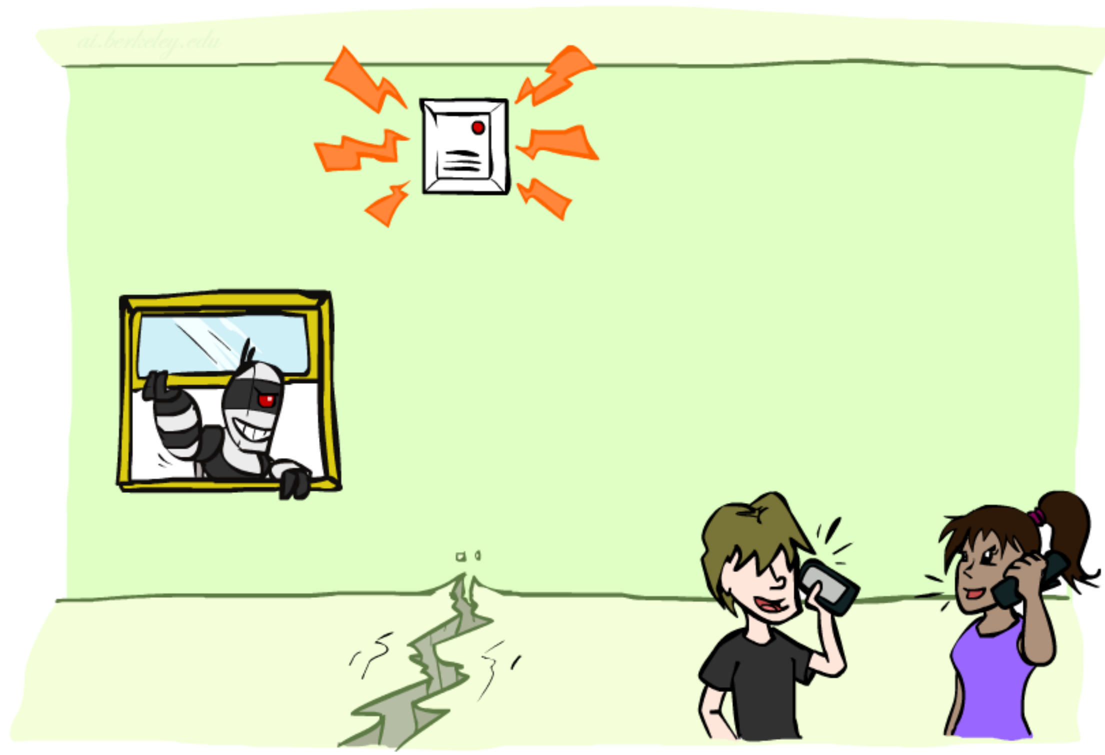
Variables

- 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!



Semantics Example

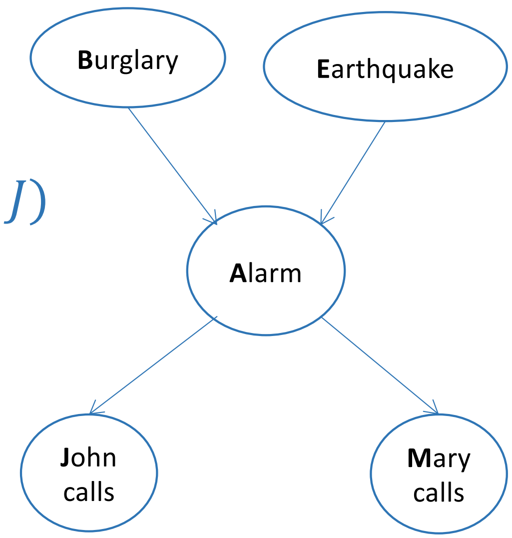
Joint distribution factorization example

Generic chain rule

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

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

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

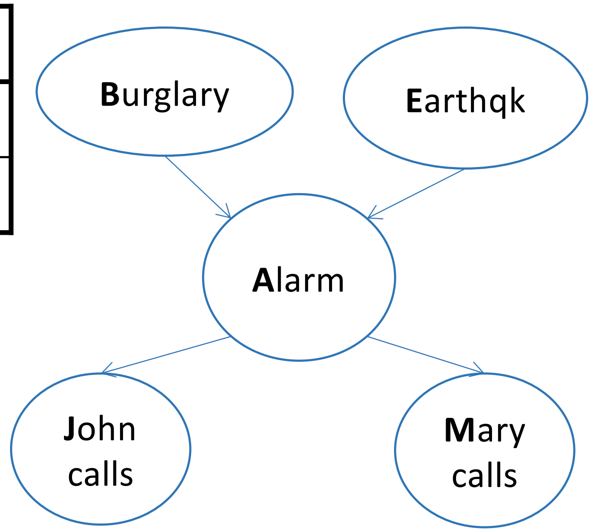


Bayes nets

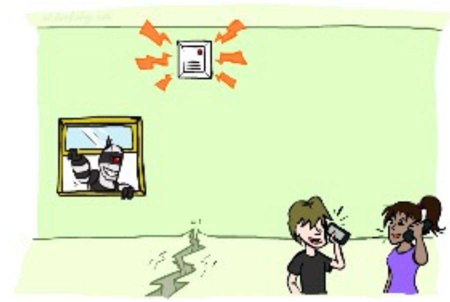
- $P(X_1 \dots X_2) = \prod_i P(X_i | Parents(X_i))$

Example: Alarm Network

B	P(B)
+b	0.001
-b	0.999



E	P(E)
+e	0.002
-e	0.998



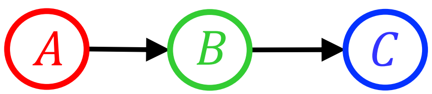
A	J	P(J|A)
+a	+j	0.9
+a	-j	0.1
-a	+j	0.05
-a	-j	0.95

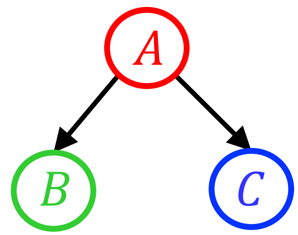
A	M	P(M|A)
+a	+m	0.7
+a	-m	0.3
-a	+m	0.01
-a	-m	0.99

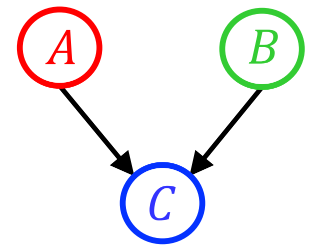
B	E	A	P(A|B,E)
+b	+e	+a	0.95
+b	+e	-a	0.05
+b	-e	+a	0.94
+b	-e	-a	0.06
-b	+e	+a	0.29
-b	+e	-a	0.71
-b	-e	+a	0.001
-b	-e	-a	0.999

Poll 2

Match the product of CPTs to the Bayes net.





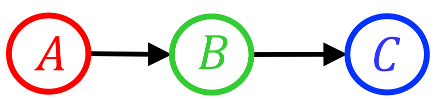


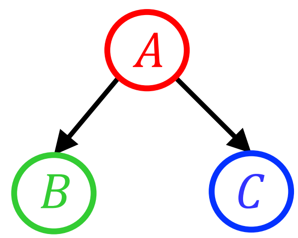
I.	$P(A)\,P(B\|A)\,P(C\|B)$	$P(A)\,P(B\|A)\,P(C\|A)$	$P(A)\,P(B)\,P(C\|A,B)$
II.	$P(A)\,P(B)\,P(C\|A,B)$	$P(A)\,P(B\|A)\,P(C\|B)$	$P(A)\,P(B\|A)\,P(C\|A)$
III.	$P(A)\,P(B\|A)\,P(C\|B)$	$P(A\|B,C)\,P(B)\,P(C)$	$P(A)\,P(B\|A)\,P(C\|B)$

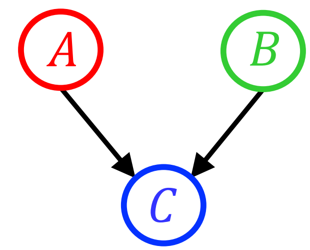
Conditional Independence Semantics

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)



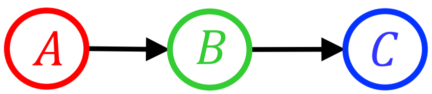




Conditional Independence Semantics

For the following Bayes nets, write the joint $P(A, B, C)$

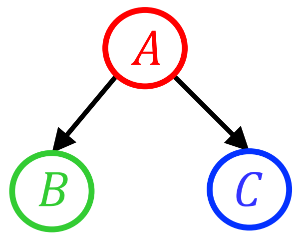
1. Using the chain rule (with top-down order A,B,C)
2. Using Bayes net semantics (product of CPTs)



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

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

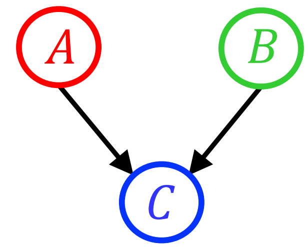
Assumption:
$P(C|A, B) = P(C|B)$
C is independent from A given B



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

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

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 *always* causes rain
  - Rain *always* causes traffic
  - High pressure *always* causes no rain
  - No rain *always* causes no traffic
  - In numbers:

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

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

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 given Y?

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

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

$$= P(z|y)$$

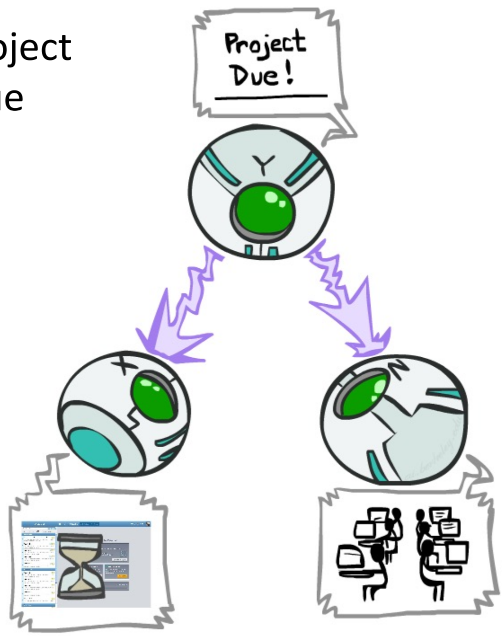
Yes!

- Evidence along the chain "blocks" the influence

Common Cause

This configuration is a "common cause"

Y: Project due



X: Forums busy

Z: Lab full

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

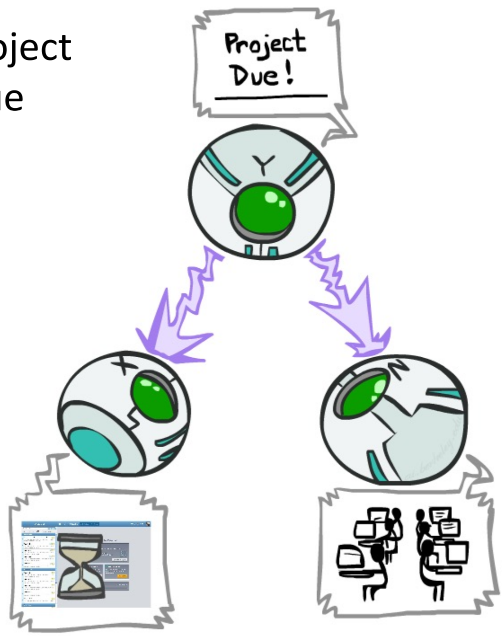
- Guaranteed X independent of Z ? *No!*
  - One example set of CPTs for which X is not independent of Z is sufficient to show this independence is not guaranteed.
  - Example:
    - Project due *always* causes both forums busy and lab full
    - In numbers:

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

Common Cause

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?

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

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

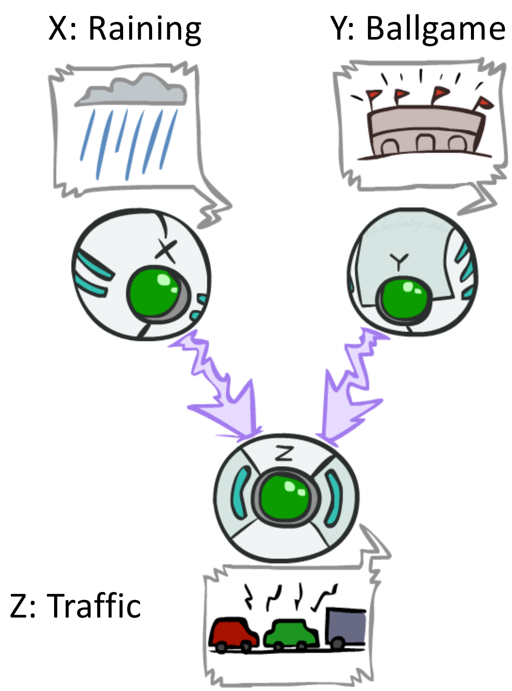
$$= P(z|y)$$

Yes!

- Observing the cause blocks influence between effects.

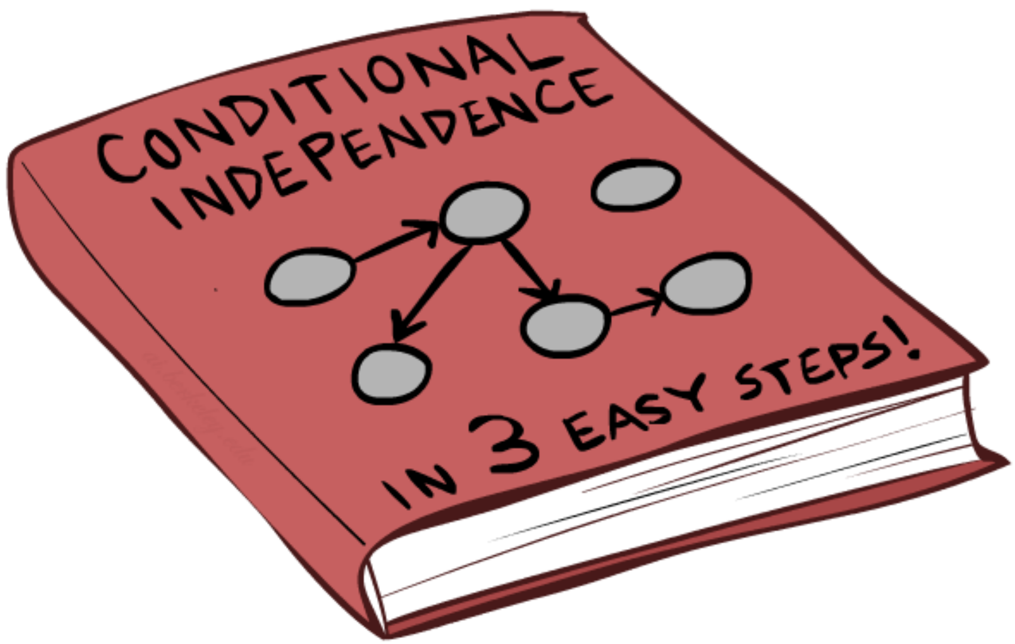
Common Effect

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



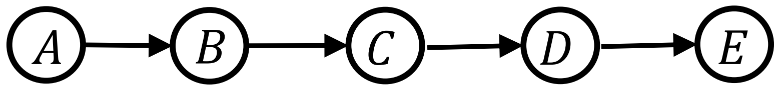
- Are X and Y independent?
  - *Yes*: the ballgame and the rain cause traffic, but they are not correlated
  - Still need to prove they must be (try it!)
- Are X and Y independent given Z?
  - *No*: seeing traffic puts the rain and the ballgame in competition as explanation.
- This is backwards from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

Bayes Net Independence

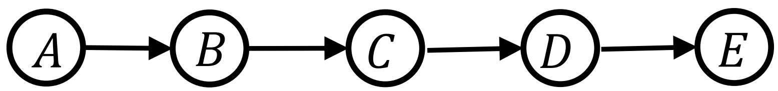


Answering Independence Questions

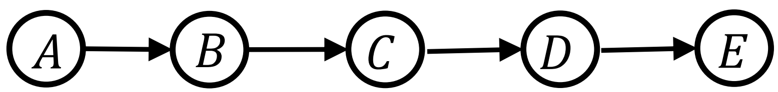
- Is A independent from E?



- Is A independent from E given C?



- Is A independent from C given E?



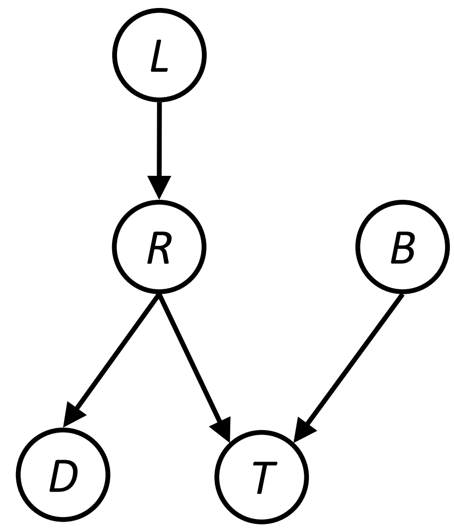
Reachability

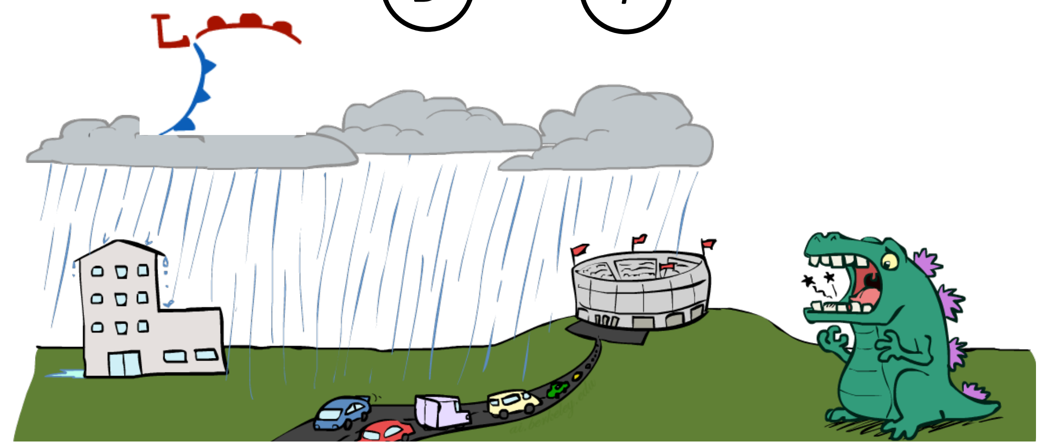
Recipe: shade evidence nodes, look for paths in the resulting graph

Attempt 1: if two nodes are connected by an undirected path not blocked by a shaded node, they are 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

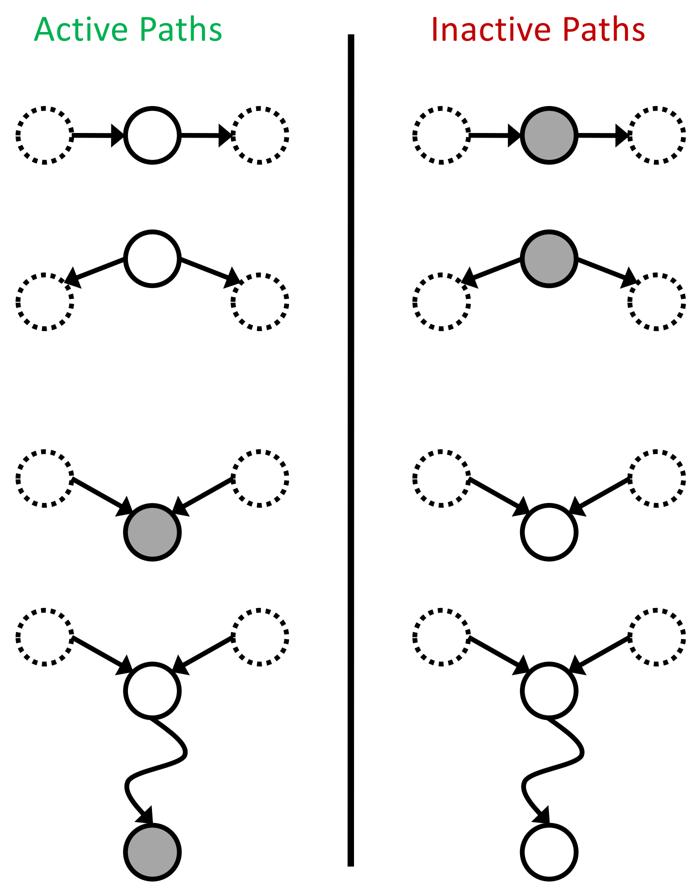
Question: Are X and Y conditionally independent given evidence variables {Z}?

- 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:

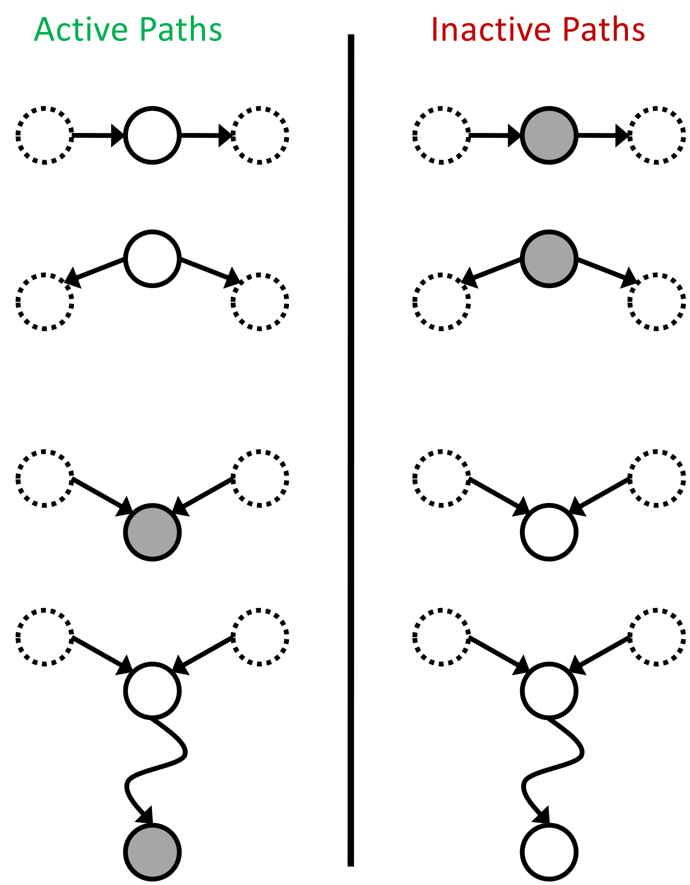
- Causal chain $A \rightarrow B \rightarrow C$ where B is unobserved (either direction)
- Common cause $A \leftarrow B \rightarrow C$ where B is unobserved
- Common effect (aka v-structure)
  $A \rightarrow B \leftarrow C$ where B *or one of its descendents* is observed

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



Bayes Ball

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

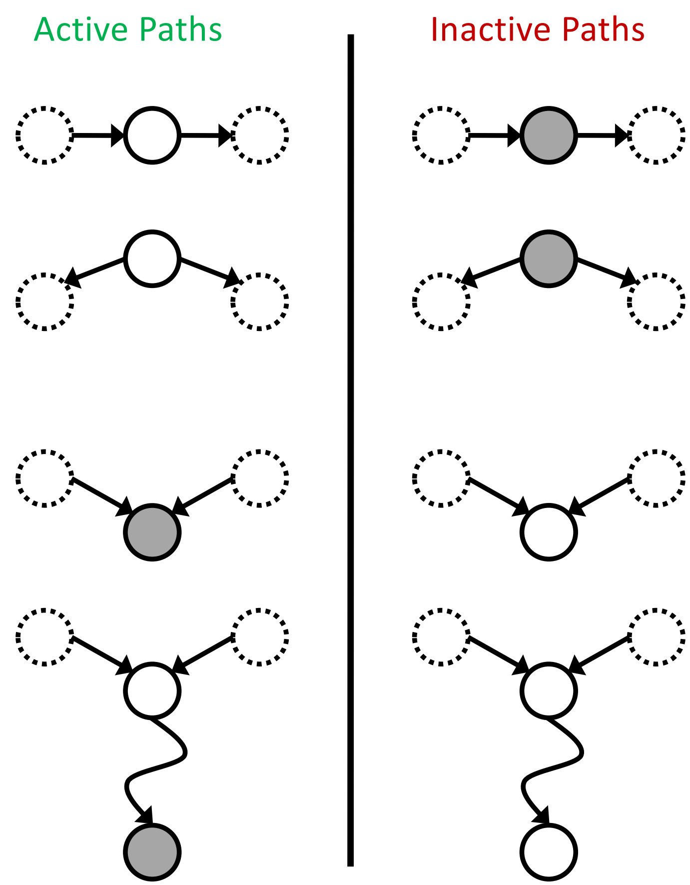


Shachter, Ross D. "Bayes-Ball: Rational Pastime (for Determining Irrelevance and Requisite Information in Belief Networks and Influence Diagrams)." *Proceedings of the Fourteenth conference on Uncertainty in Artificial Intelligence.* 1998.

Bayes Ball

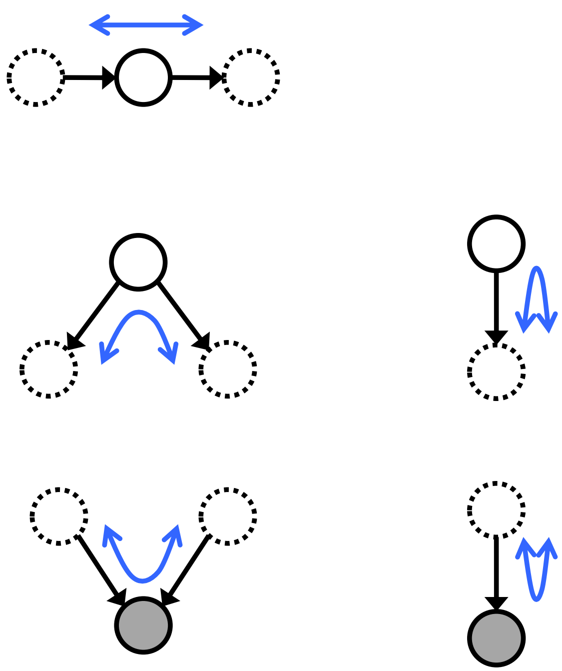
Question: Are X and Y conditionally independent given evidence variables {Z}?

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

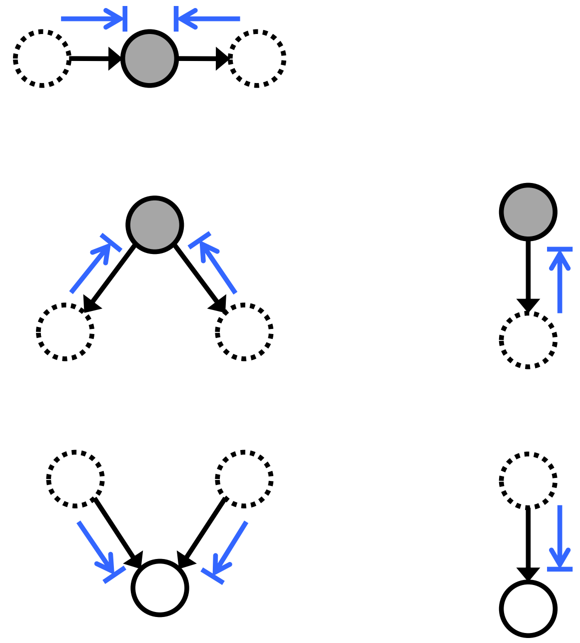


Bayes Ball

Active Paths

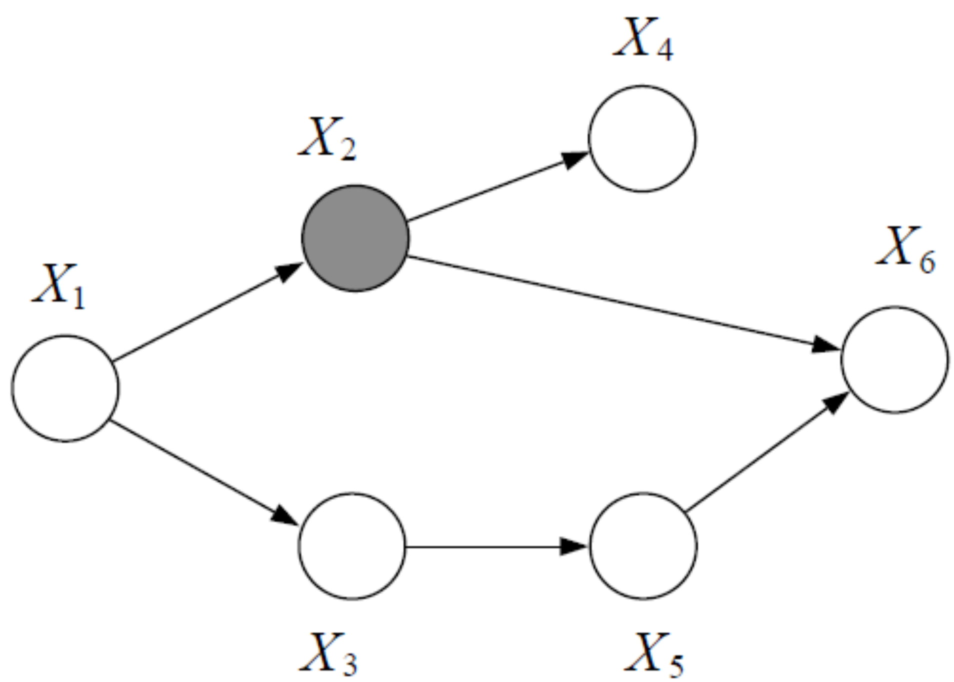


Inactive Paths



Poll 3

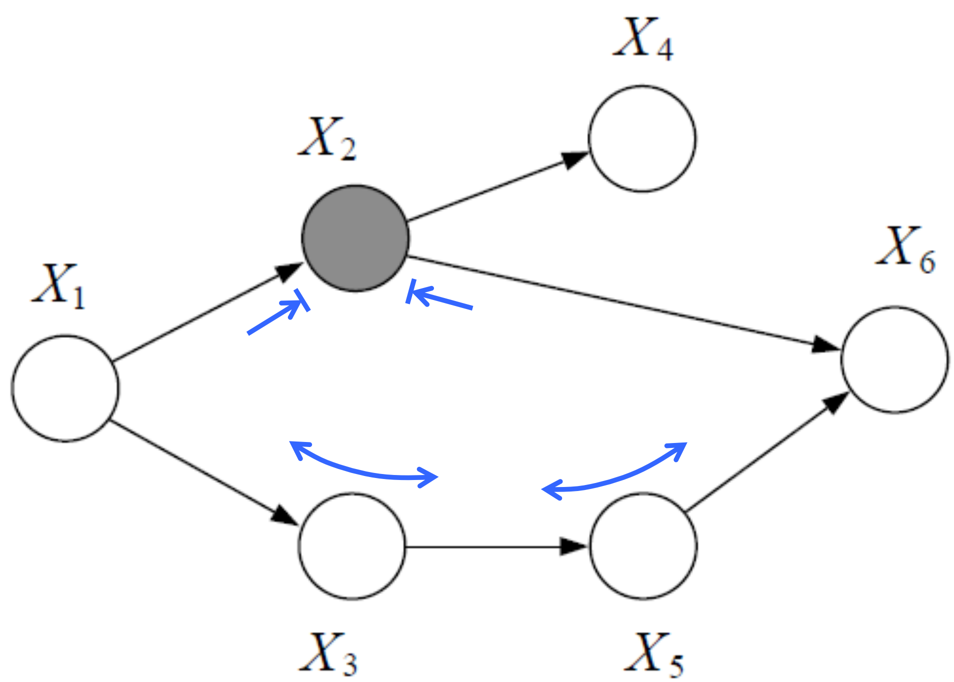
Is X_1 independent from X_6 given X_2?



Poll 3

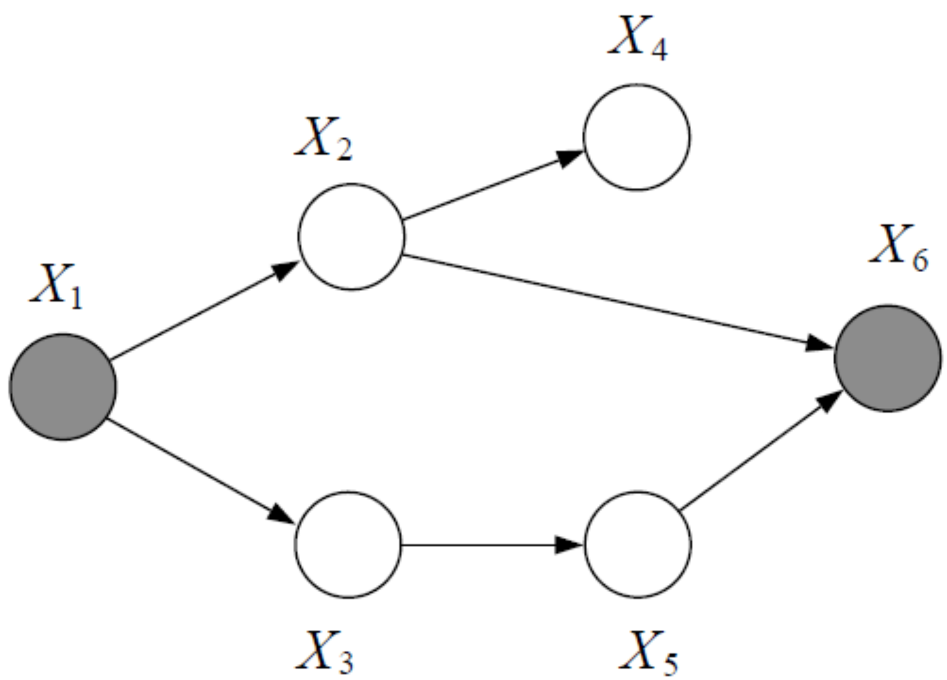
Is X_1 independent from X_6 given X_2?

No, the Bayes ball can travel through X_3 and X_5.



Poll 4

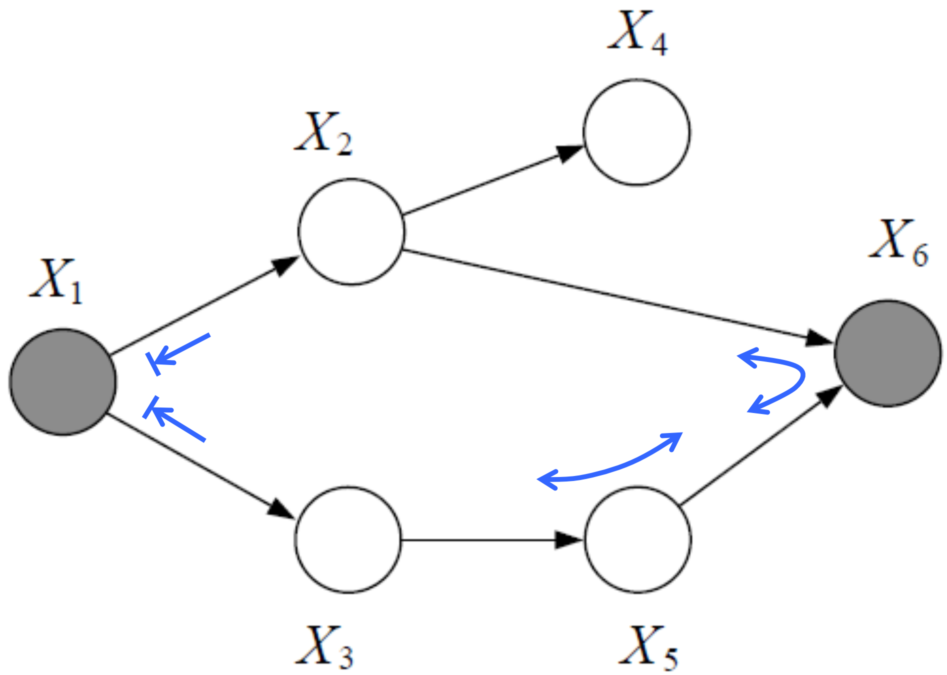
Is X_2 independent from X_3 given X_1 and X_6?



Poll 4

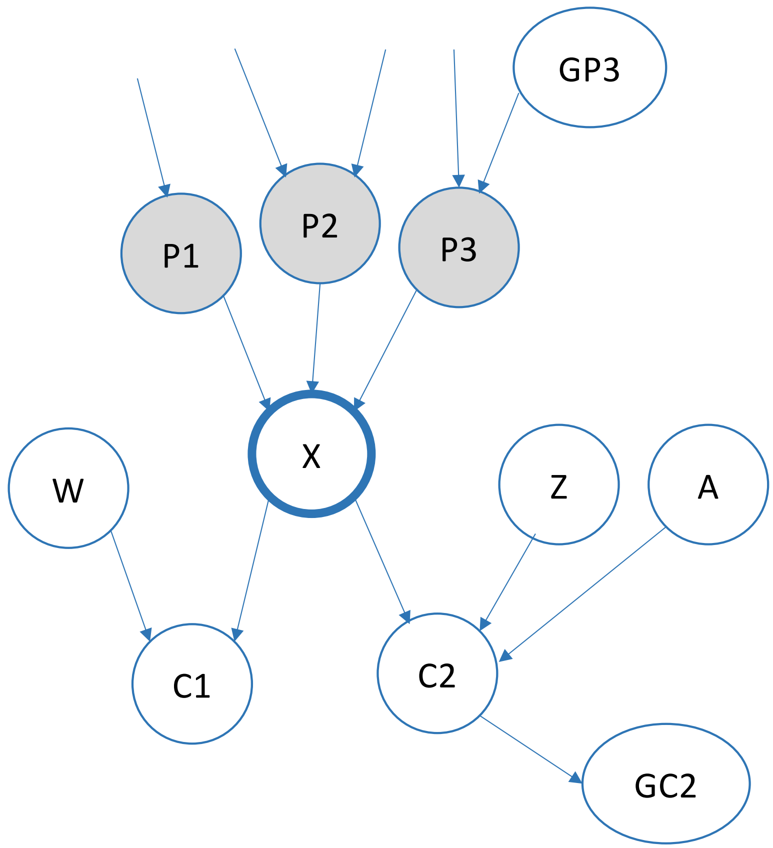
Is X_2 independent from X_3 given X_1 and X_6?

No, the Bayes ball can travel through X_5 and X_6.



Conditional Independence Semantics

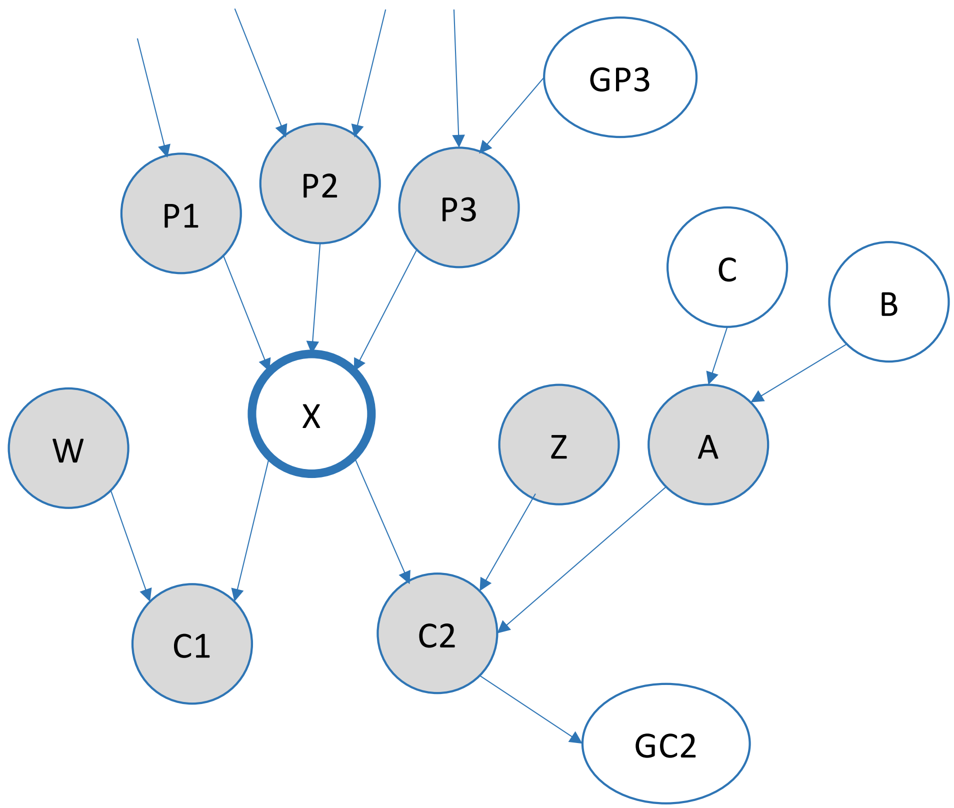
Every variable is conditionally independent of its non-descendants given its parents



Markov blanket

A variable's Markov blanket consists of parents, children, children's other parents

Every variable is conditionally independent of all other variables given its Markov blanket

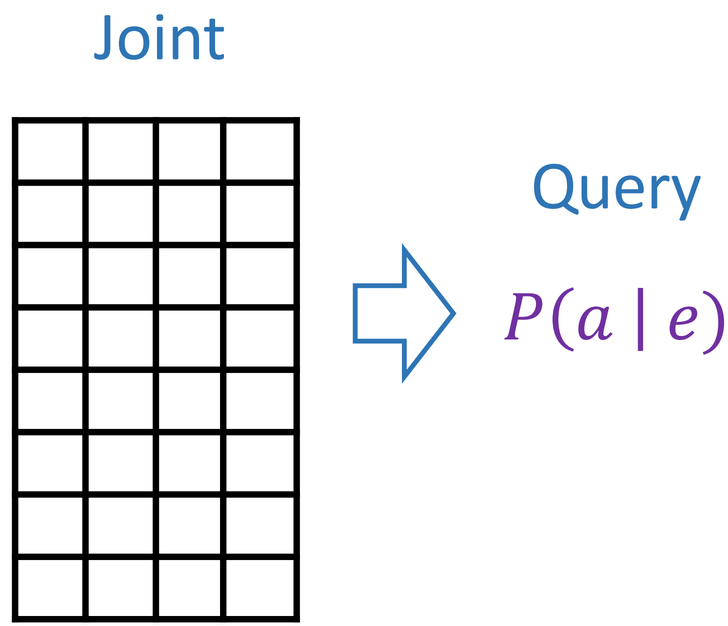


Answer Any Query from Joint Distribution

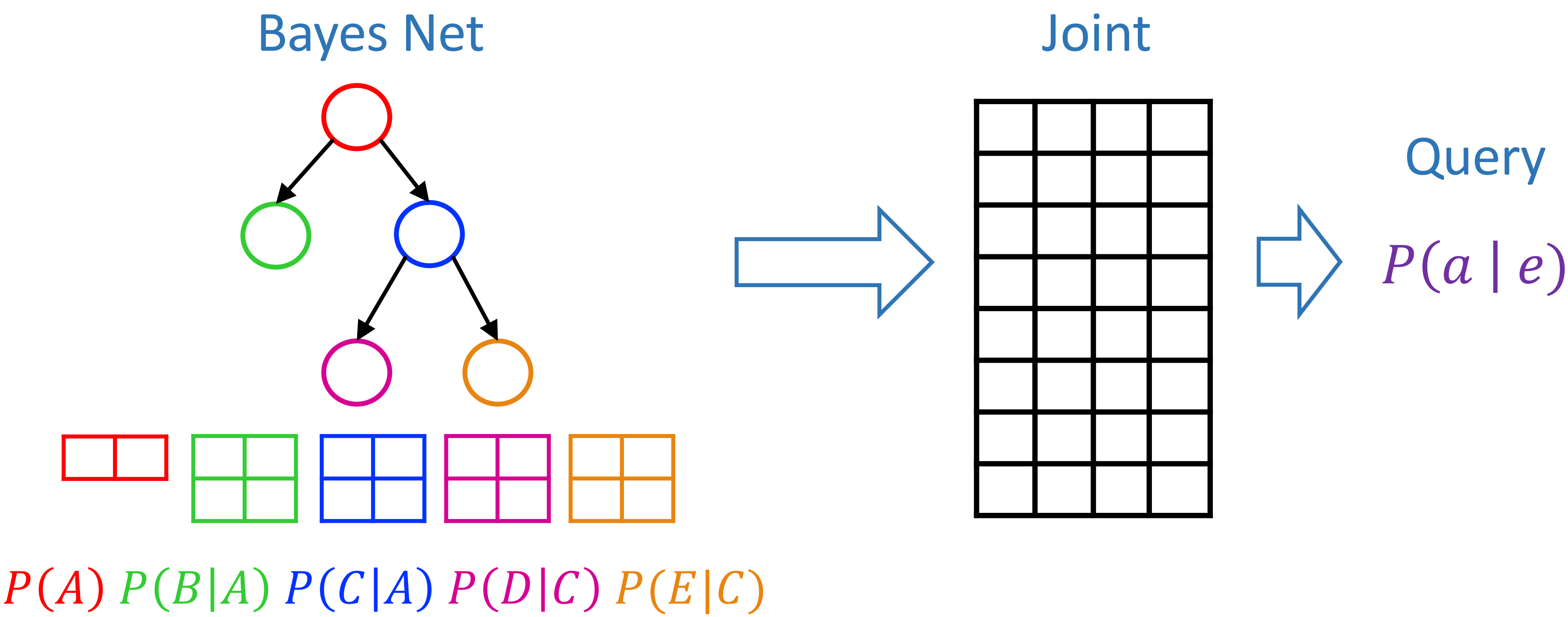
Joint distributions are the best!

Problems with joints

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



Answer Any Query from Bayes Net



Next: Answer Any Query from Bayes Net