Artificial Intelligence

CS 165A

Oct 20, 2020

Instructor: Prof. Yu-Xiang Wang

Today

- → Factorization and conditional independence
- → Bayesian Network Examples
- → Conditional Independence

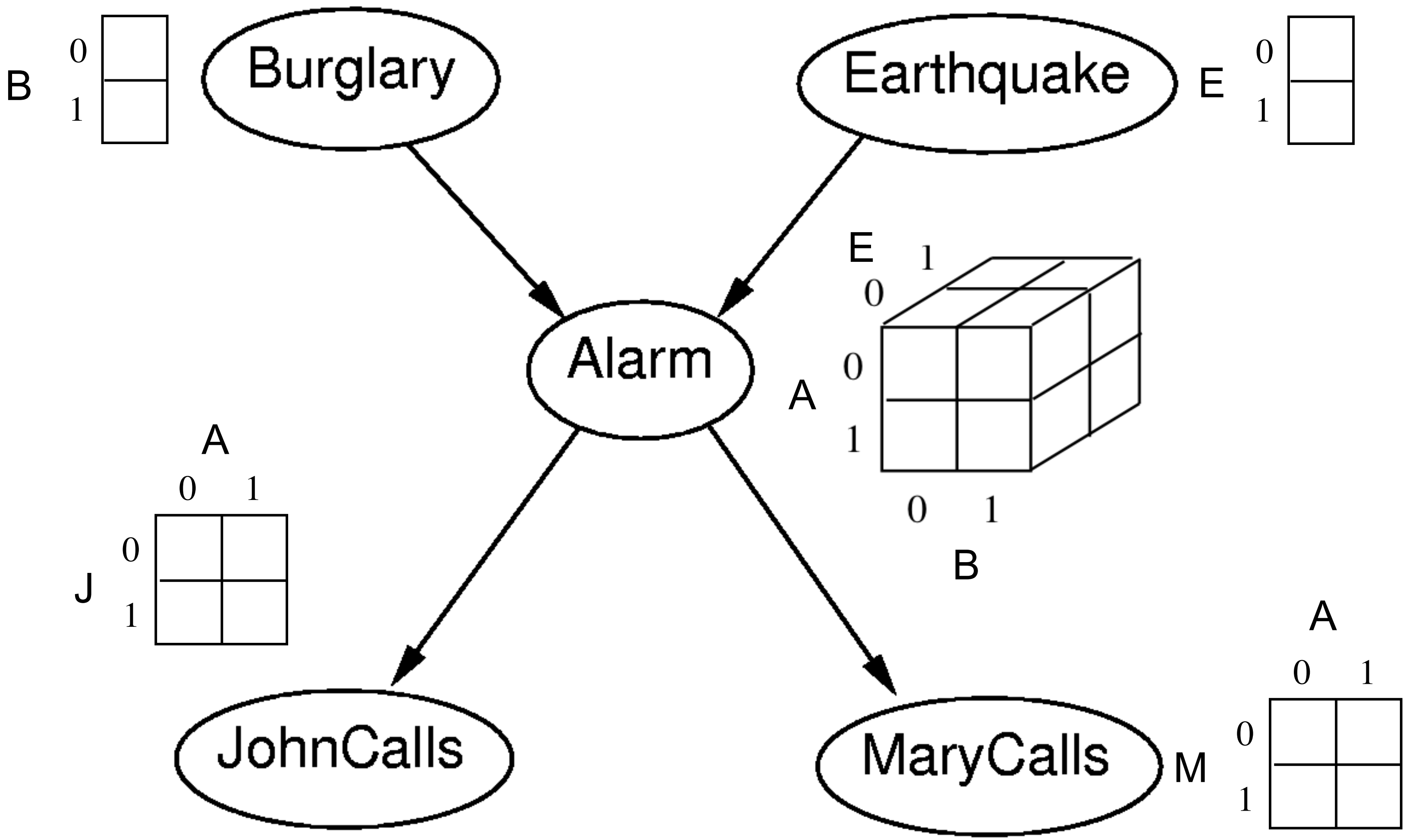
Recap: Example: Modelling with Belief Net

I'm at work and my neighbor John called to say my home alarm is ringing, but my neighbor Mary didn't call. The alarm is sometimes triggered by minor earthquakes. Was there a burglar at my house?

- Random (boolean) variables:
  - JohnCalls, MaryCalls, Earthquake, Burglar, Alarm
- The belief net shows the causal links
- This defines the joint probability
  - P(JohnCalls, MaryCalls, Earthquake, Burglar, Alarm)
- What do we want to know?

$P(B \mid J, \neg M)$

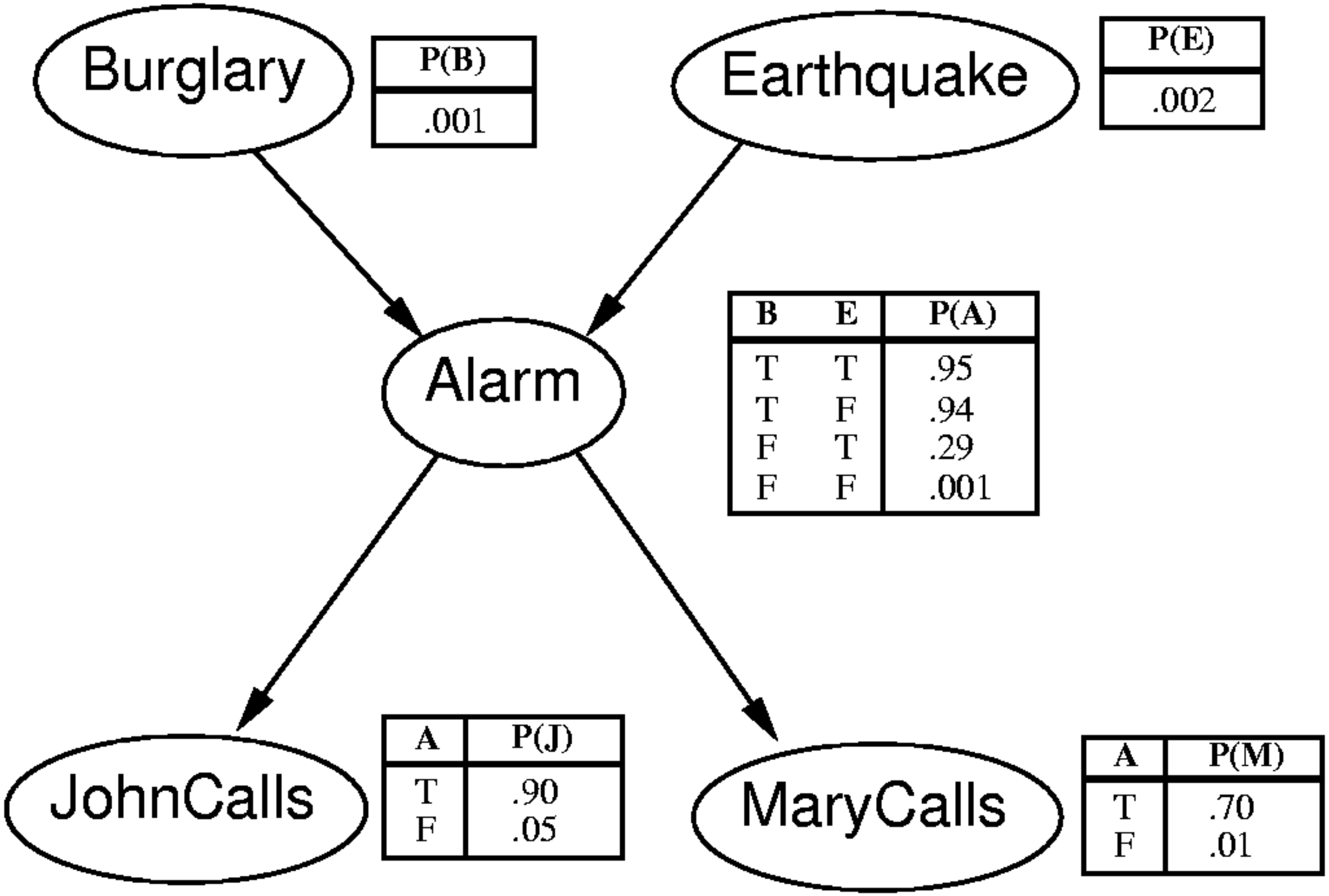
Recap: What are the CPTs? What are their dimensions?



Question: How to fill values into these CPTs?

Ans: Specify by hands. Learn from data (e.g., MLE).

Recap: Example



P(B)
.001

P(E)
.002

B	E	P(A)
T	T	.95
T	F	.94
F	T	.29
F	F	.001

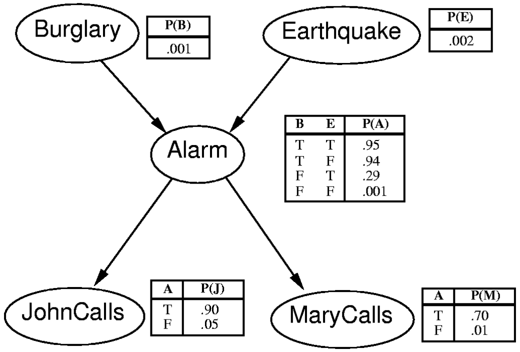
A	P(J)
T	.90
F	.05

A	P(M)
T	.70
F	.01

Joint probability? P(J, ¬M, A, B, ¬E)?

This lecture

- Continue with the above example
  - Probabilistic inference via marginalization
- Conditional independence
- Reading off Conditional Independences from a Bayesian Network
  - d-separation
  - Bayes Ball algorithm
  - Markov Blanket



Calculate $P(J, \neg M, A, B, \neg E)$

Read the joint pf from the graph:

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

Plug in the desired values:

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

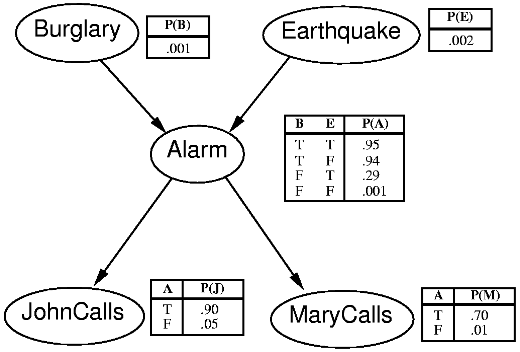
$$= 0.001 * 0.998 * 0.94 * 0.9 * 0.3$$

$$= 0.0002532924$$

How about $P(B \mid J, \neg M)$?

Remember, this means $P(B=\text{true} \mid J=\text{true}, M=\text{false})$

Calculate P(B | J, ¬M)



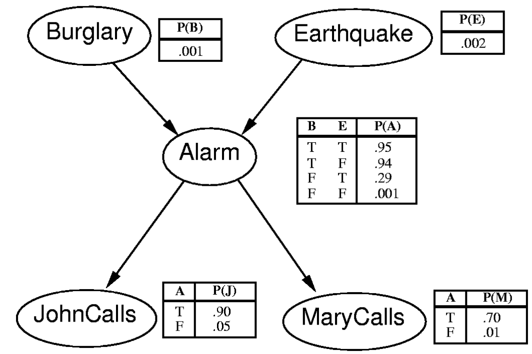
$$P(B \mid J, \neg M) = \frac{P(B, J, \neg M)}{P(J, \neg M)}$$

By marginalization:

$$= \frac{\sum_i \sum_j P(J, \neg M, A_i, B, E_j)}{\sum_i \sum_j \sum_k P(J, \neg M, A_i, B_j, E_k)}$$

$$= \frac{\sum_i \sum_j P(B)P(E_j)P(A_i \mid B, E_j)P(J \mid A_i)P(\neg M \mid A_i)}{\sum_i \sum_j \sum_k P(B_j)P(E_k)P(A_i \mid B_j, E_k)P(J \mid A_i)P(\neg M \mid A_i)}$$

Variable elimination algorithm



$$P(B \mid J, \neg M) = \frac{P(B, J, \neg M)}{P(J, \neg M)}$$

$$= \frac{\sum_i \sum_j P(B)P(E_j)P(A_i \mid B, E_j)P(J \mid A_i)P(\neg M \mid A_i)}{\sum_i \sum_j \sum_k P(B_j)P(E_k)P(A_i \mid B_j, E_k)P(J \mid A_i)P(\neg M \mid A_i)}$$

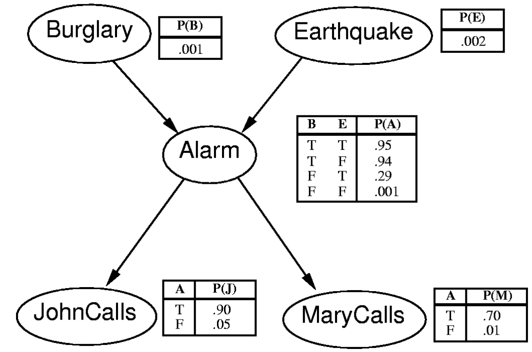
*Exchange the order of summation and product

Quick checkpoint

- Bayesian Network as a modelling tool
- By inspecting the cause-effect relationships, we can draw directed edges based on our domain knowledge
- The product of the CPTs give the joint distribution
  - We can calculate P(A | B) for any A and B
  - The factorization makes it computationally more tractable

What else can we get?

Example: Conditional Independence



- Conditional independence is seen here
  - P(JohnCalls | MaryCalls, Alarm, Earthquake, Burglary) = P(JohnCalls | Alarm)
  - So JohnCalls is independent of MaryCalls, Earthquake, and Burglary, given Alarm

- Does this mean that an earthquake or a burglary do not influence whether or not John calls?
  - No, but the influence is already accounted for in the Alarm variable
  - JohnCalls is conditionally independent of Earthquake, but not marginally independent of it

***This conclusion is independent to values of CPTs!**

Question

If X and Y are independent, are they therefore independent given any variable(s)?

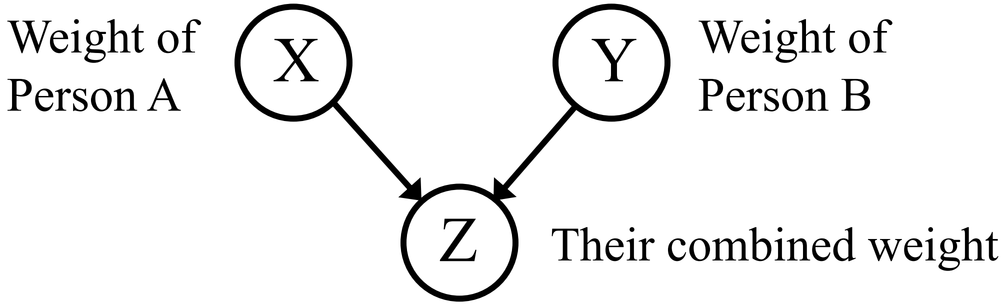
I.e., if $P(X, Y) = P(X)\,P(Y)$ [i.e., if $P(X|Y) = P(X)$], can we conclude that
$P(X \mid Y, Z) = P(X \mid Z)$?

Question

If X and Y are independent, are they therefore independent given any variable(s)?

I.e., if $P(X, Y) = P(X)\,P(Y)$ [i.e., if $P(X|Y) = P(X)$], can we conclude that
$P(X \mid Y, Z) = P(X \mid Z)$?

The answer is **no**, and here's a counter example:



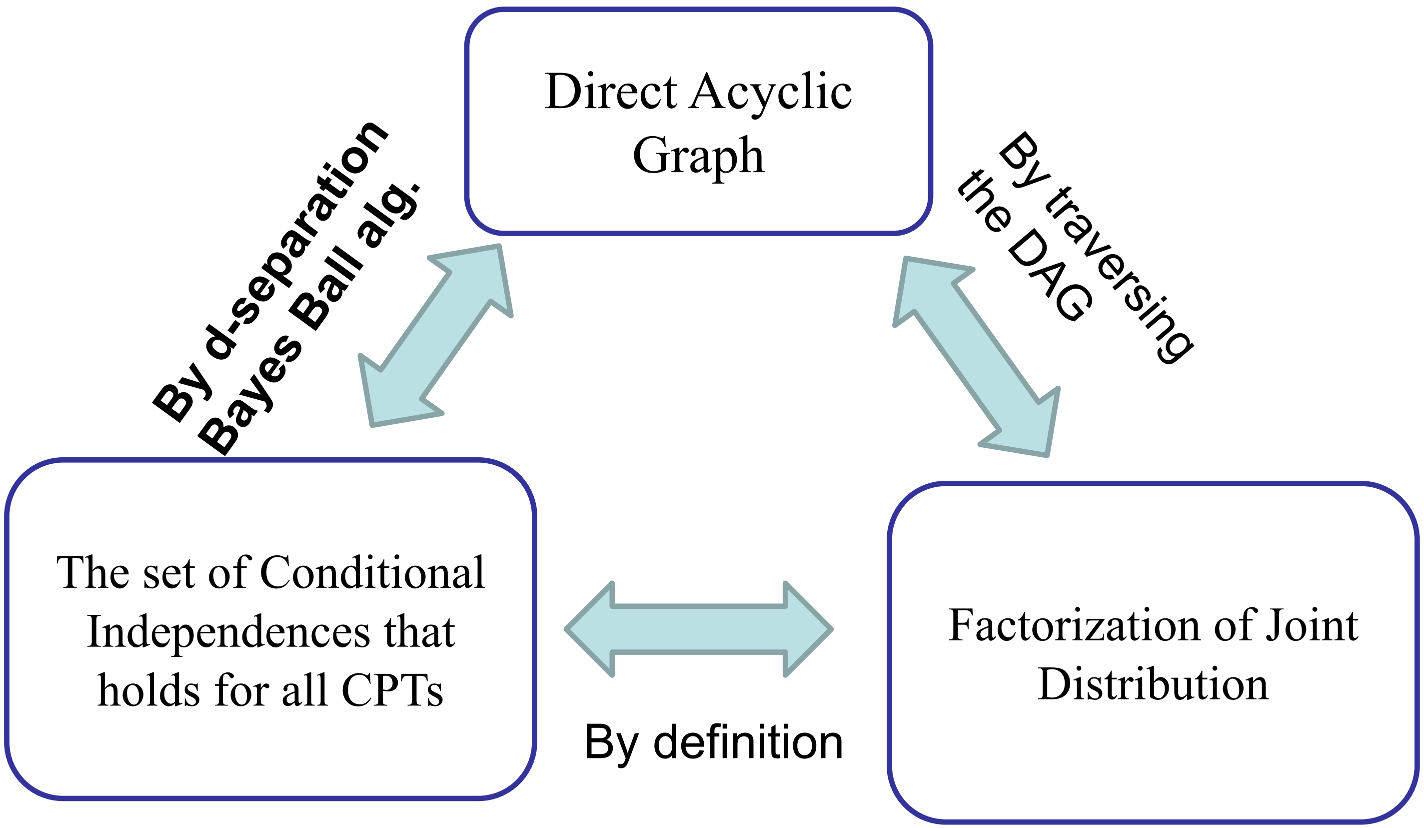
$$P(X \mid Y) = P(X)$$
$$P(X \mid Y, Z) \neq P(X \mid Z)$$

Note: Even though Z is a deterministic function of X and Y, it is still a random variable with a probability distribution

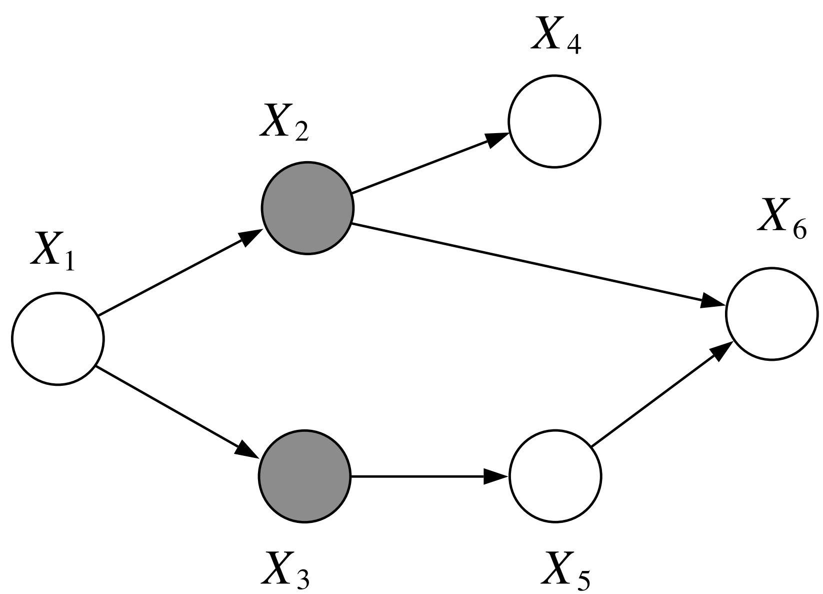
*Again: This conclusion is independent to values of CPTs!

Big question: Is there a general way that we can answer questions about conditional independences by just inspecting the graphs?

- Turns out the answer is "Yes!"



Intuition: the graph and the edges controls the information flow, if there is no path that the information can flow from one-node to another, we say these two nodes are independent..

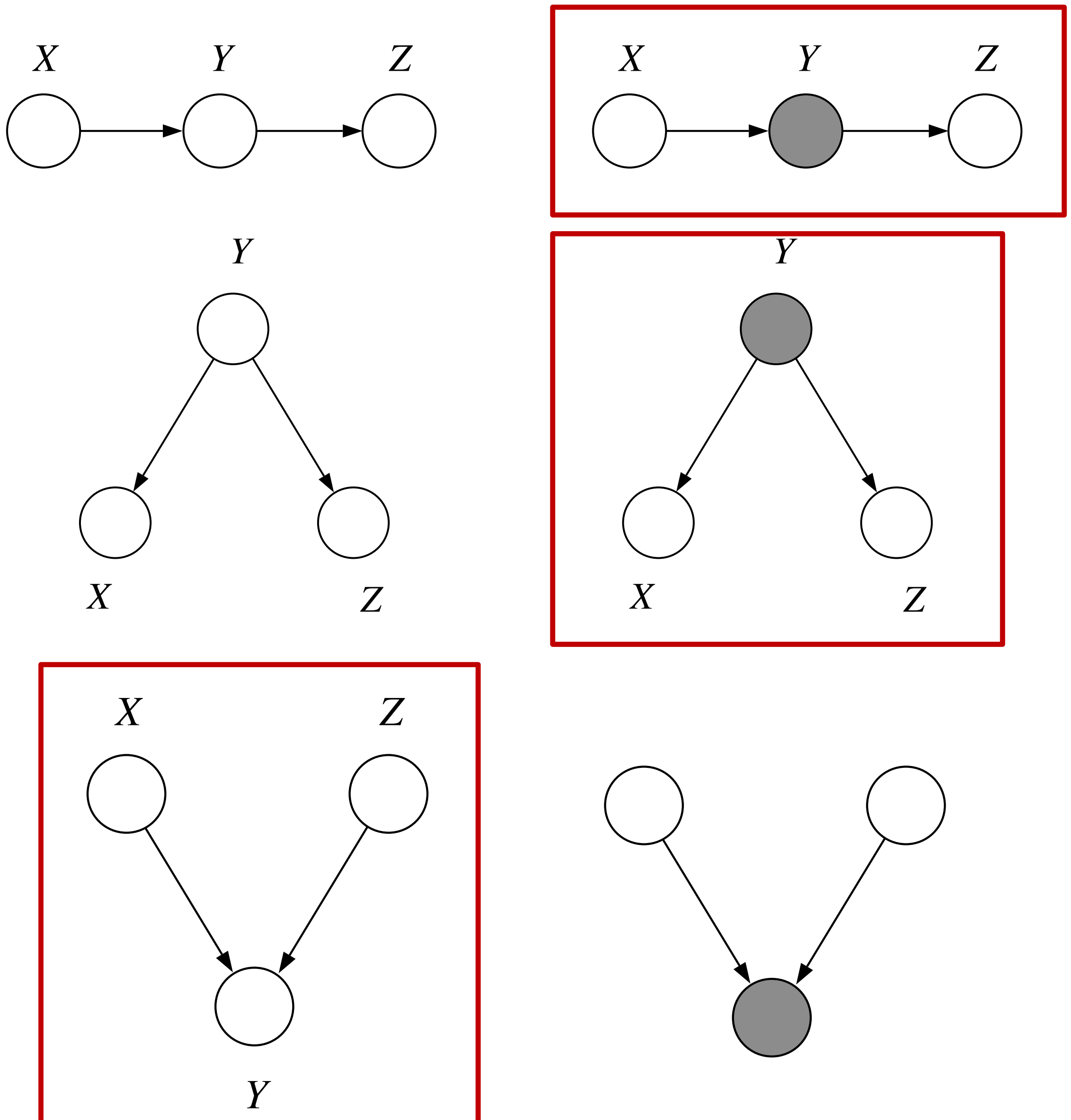


"Shading" denotes "observing" or "conditioning on" that variables.

Figure 2.3: The nodes X_2 and X_3 separate X_1 from X_6.

From Chapter 2 of Jordan book "Introduction to Probabilistic Graphical Models"

d-separation in three canonical graphs



$\mathbf{X \perp Z \mid Y}$

"**Chain:** X and Z are d-separated by the observation of Y."

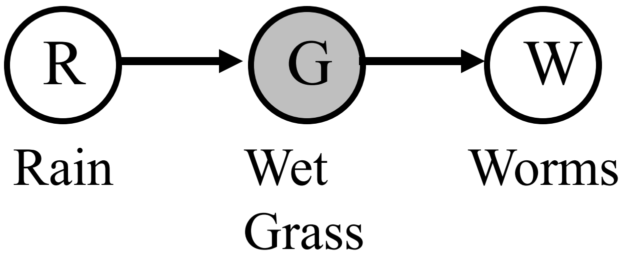
$\mathbf{X \perp Z \mid Y}$

"**Fork:** X and Z are d-separated by the observation of Y."

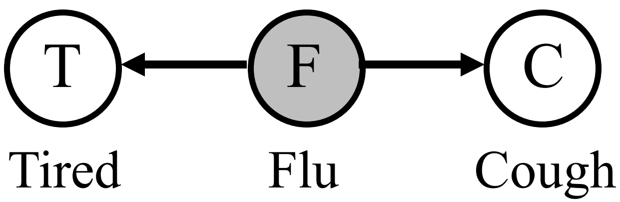
$\mathbf{X \perp Z}$

"**Collider:** X and Z are d-separated by NOT observing Y **nor any descendants** of Y."

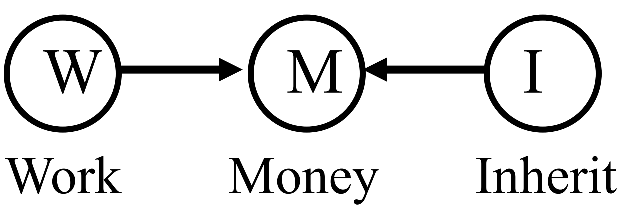
Examples



$$P(W \mid R, G) = P(W \mid G)$$



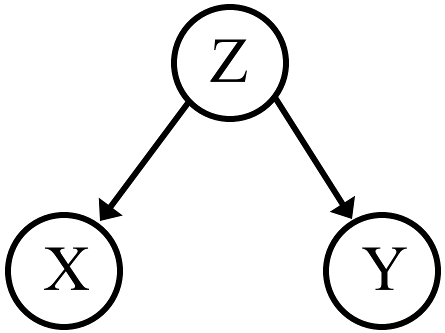
$$P(T \mid C, F) = P(T \mid F)$$



$$P(W \mid I, M) \neq P(W \mid M)$$

$$P(W \mid I) = P(W)$$

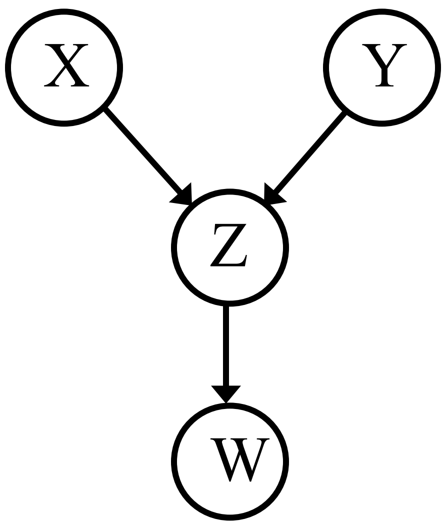
Examples



X – wet grass
Y – rainbow
Z – rain

P(X, Y) ≠ P(X) P(Y)
P(X | Y, Z) = P(X | Z)

Are X and Y ind.? Are X and Y cond. ind. given…?



X – rain
Y – sprinkler
Z – wet grass
W – worms

P(X, Y) = P(X) P(Y)
P(X | Y, Z) ≠ P(X | Z)
P(X | Y, W) ≠ P(X | W)

The Bayes Ball algorithm

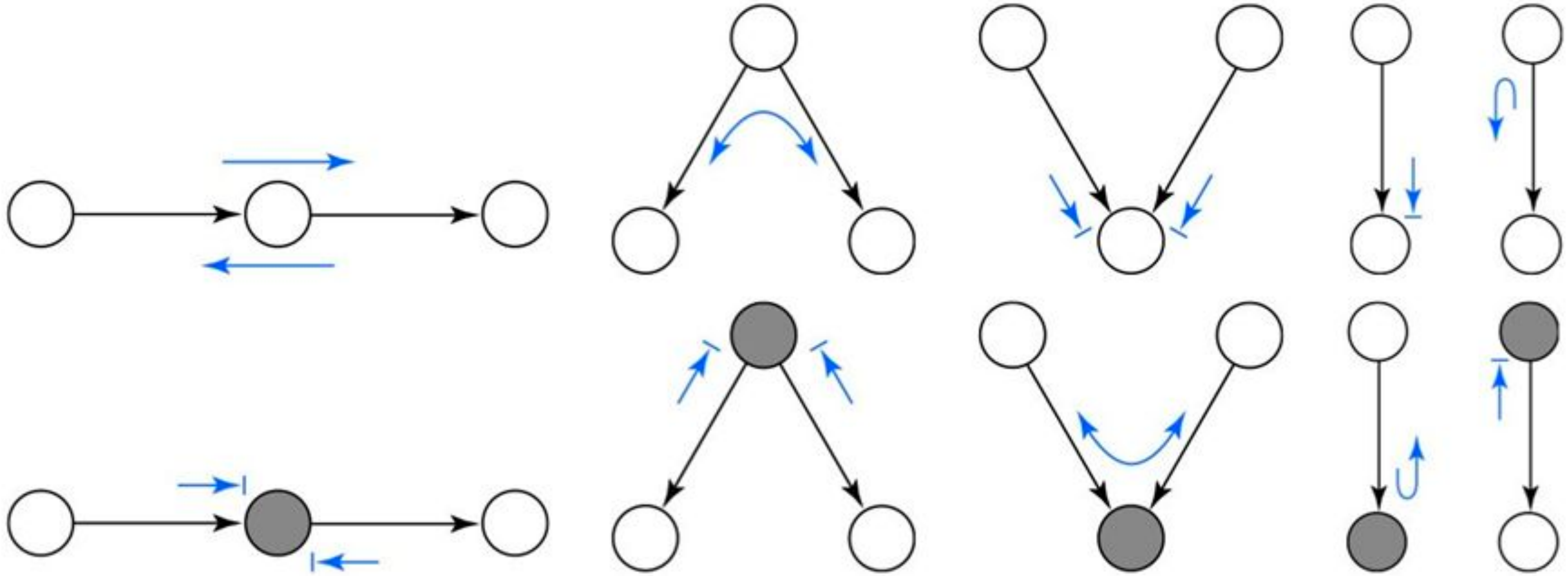
- Let **X**, **Y**, **Z** be "***groups***" of nodes / set / subgraphs.

- Shade nodes in **Y**
- Place a "ball" at each node in **X**
- Bounce balls around the graph according to **rules**

- If no ball reaches any node in **Z**, then declare

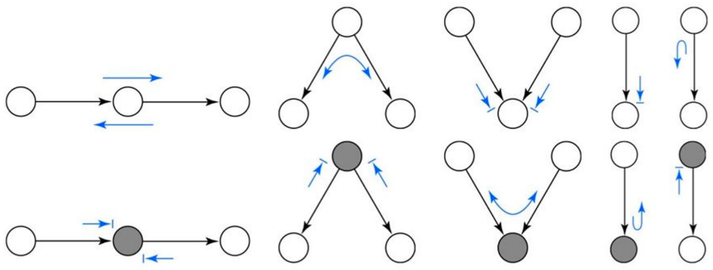
$$\mathbf{X} \perp \mathbf{Z} \mid \mathbf{Y}$$

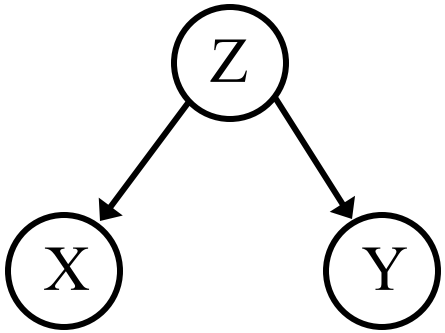
The Ten Rules of Bayes Ball Algorithm



Please read [Jordan PGM Ch. 2.1]
to learn more about the Bayes Ball algorithm

Examples (revisited using Bayes-ball alg)

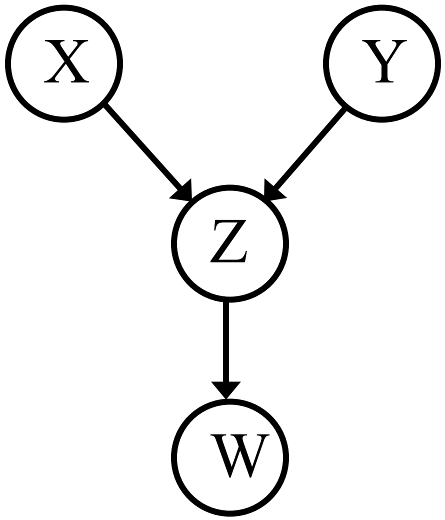




X – wet grass
Y – rainbow
Z – rain

$P(X, Y) \neq P(X)\,P(Y)$
$P(X \mid Y, Z) = P(X \mid Z)$

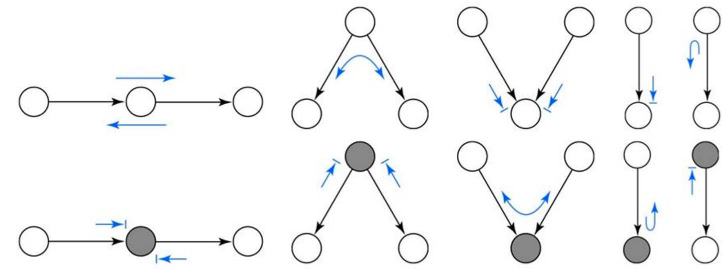
Are X and Y ind.? Are X and Y cond. ind. given…?



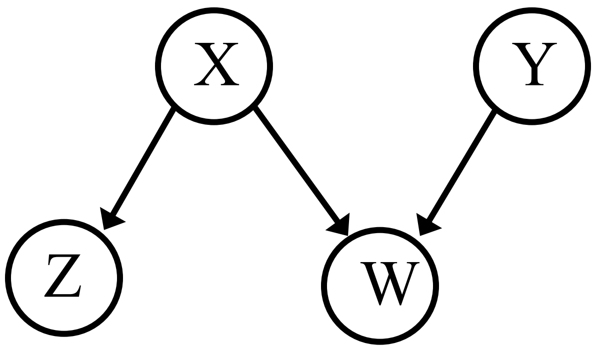
X – rain
Y – sprinkler
Z – wet grass
W – worms

$P(X, Y) = P(X)\,P(Y)$
$P(X \mid Y, Z) \neq P(X \mid Z)$
$P(X \mid Y, W) \neq P(X \mid W)$

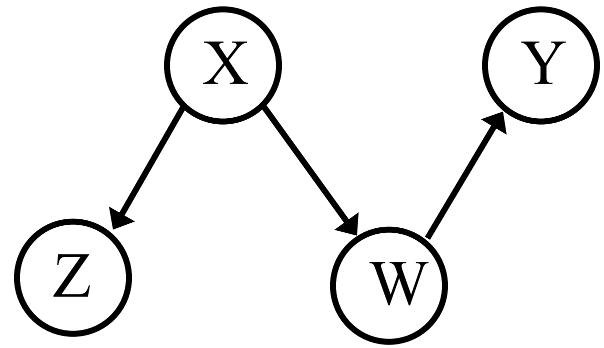
Examples (3 min work)



Are X and Y independent?
Are X and Y conditionally independent given Z?

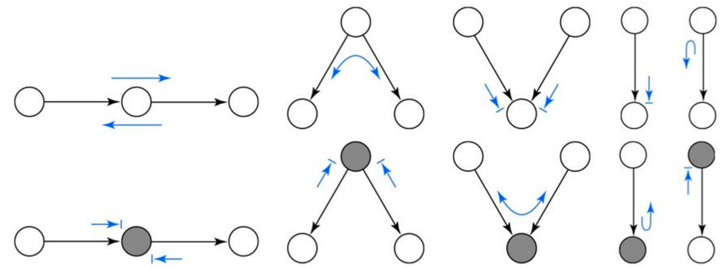


X – rain
Y – sprinkler
Z – rainbow
W – wet grass



X – rain
Y – sprinkler
Z – rainbow
W – wet grass

Conditional Independence



- Where are conditional independences here?

Radio and Ignition, given Battery?

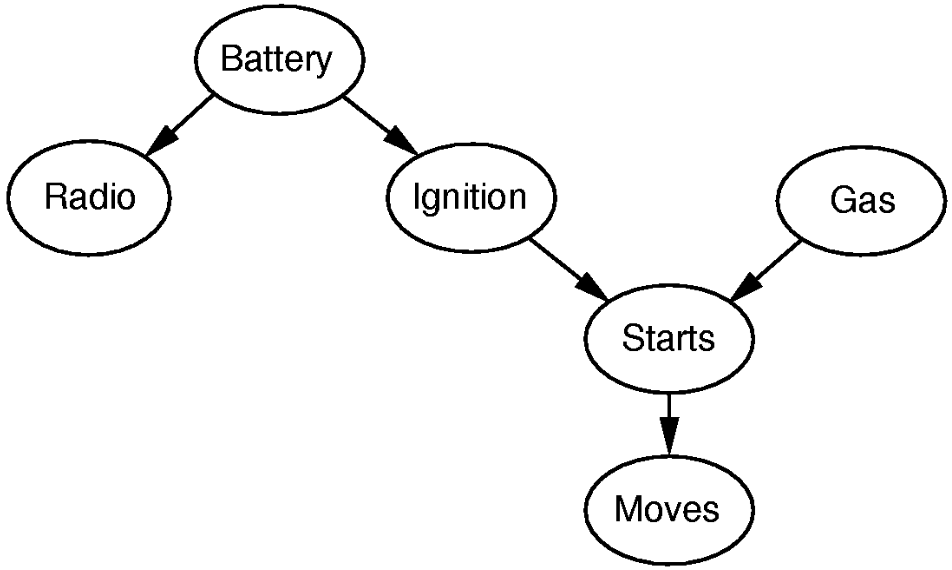
Radio and Starts, given Ignition?

Gas and Radio, given Battery?

Gas and Radio, given Starts?

Gas and Radio, given nil?

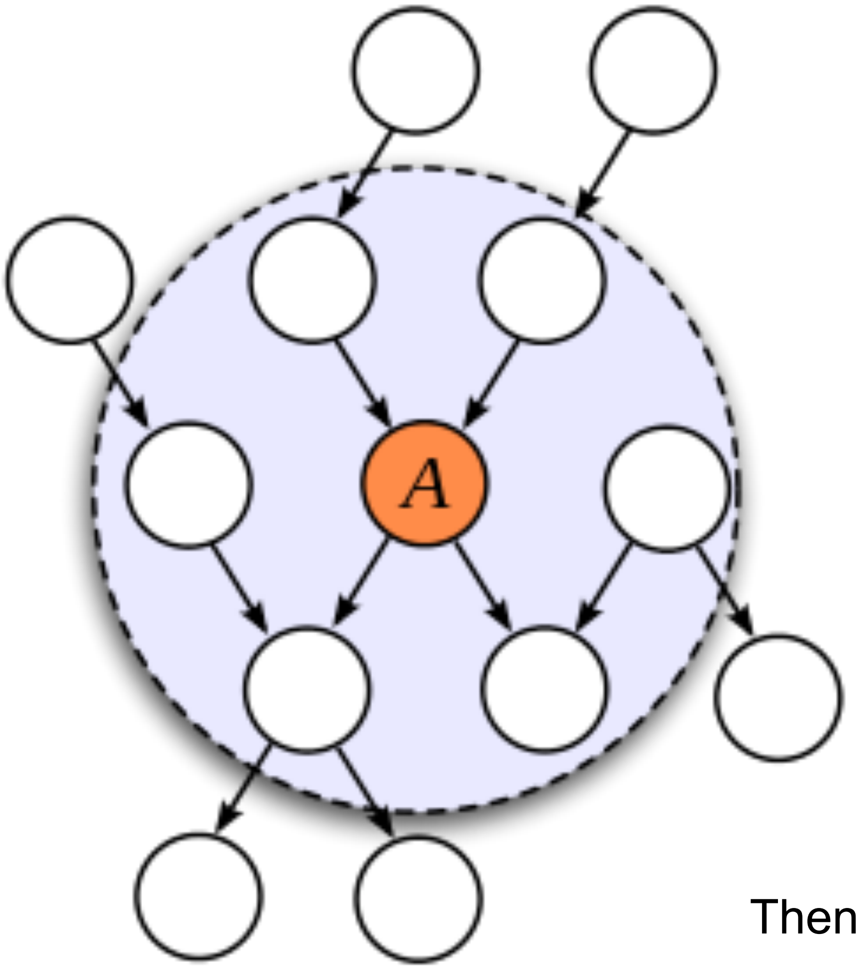
Gas and Battery, given Moves?



Quick checkpoint

- Reading conditional independences from the DAG itself.
- d-separation
  - Three canonical graphs: Chain, Fork, Collider
- Bayes ball algorithm for determining whether $\mathbf{X} \perp \mathbf{Z} \mid \mathbf{Y}$
  - Bounce the ball from any node in X by following the ten rules
  - If any ball reaches any node in Z, then return "False"
  - Otherwise, return "True"

An alternative view: Markov Blankets



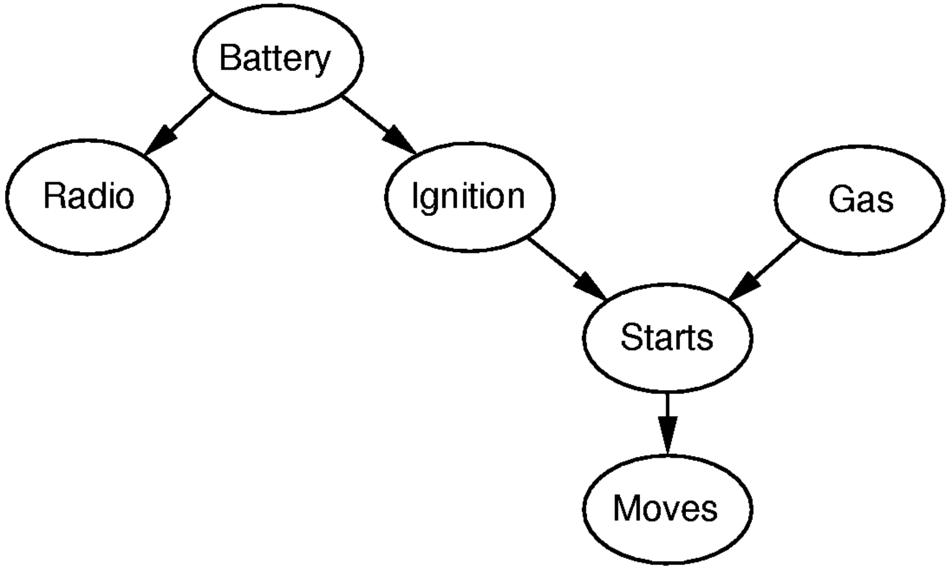
1. Parents

2. Children

3. Children's other parents

Then A is d-separated from everything else.

Example: Markov Blankets



- Question: What is the Markov Blanket of …
  - "Ignition":
  - "Starts":

Why are conditional independences important?

- Helps the developer (or the user) verify the graph structure
  - Are these variables really independent?
  - Do I need more/fewer edges in the graphical model?

- Statistical tests for (Conditional) Independence
  - Hilbert-Schmidt Independence Criterion (not covered)

- Hints on computational efficiencies

- Shows that you understand BNs…

Inference in Bayesian networks

- We've seen how to compute any probability from the Bayesian network
  - This is ***probabilistic inference***
    - P(Query | Evidence)
  - Since we know the joint probability, we can calculate anything via marginalization
    - P(Query, Evidence) / P(Evidence)

- However, things are usually not as simple as this
  - Structure is large or very complicated
  - **Calculation by marginalization is often intractable**
  - Bayesian inference is NP hard in space and time!!
  - (Details in AIMA Ch. 14.4 (Ch 13.4 in the Fourth Edition))

Inference in Bayesian networks (cont.)

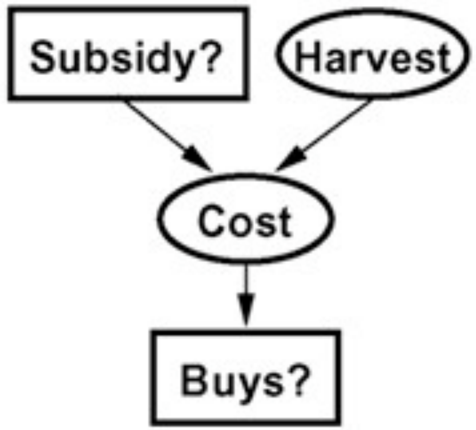
- So in all but the most simple BNs, probabilistic inference is not really done just by marginalization

- Instead, there are practical algorithms for doing approximate probabilistic inference
  - Recall a similar argument in surrogate losses in ML

- Markov Chain Monte Carlo, Message Passing / Loopy Belief Propagation
  - Active area of research!

- We won't cover these probabilistic inference algorithms though…. (Read Ch. 14.5 in the AIMA book (Ch 13.5 in the Fourth Edition))

One more thing: Continuous Variables?



You will see GMM in the discussion class.

- Dimension check: What are the shapes of the CPTs?
- Discretize? Very large CPT..
- Usually, we parametrize the conditional distribution.
  - e.g., P(Cost | Harvest) = Poisson(θ^T Harvest)

Summary of the today

- Encode knowledge / structures using a DAG
- How to check conditional independence algebraically by the factorizations?
- How to read off conditional independences from a DAG
  - d-separation, Bayes Ball algorithm, Markov Blanket
- Remarks on BN inferences and continuous variables

(**More examples in the discussion: Hidden Markov Models, AIMA 15.3 or 14.3 in the 4th Edition**)

Additional resources about PGM

- Recommended: Ch.2 Jordan book. AIMA Ch. 13-14.

- More readings:
  - Koller's PGM book: https://www.amazon.com/Probabilistic-Graphical-Models-Daphne-Koller/dp/B007YXTT12
  - Probabilistic programming: http://probabilistic-programming.org/wiki/Home

- Software for PGMs and modeling and inference:
  - Stan: https://mc-stan.org/
  - JAGS: http://mcmc-jags.sourceforge.net/

Upcoming lectures

- Oct 22: Problem solving by search
- Oct 27: Search algorithms
- Oct 29: Minimax search and game playing
- Nov 3: Midterm review. HW2 Due.

- Recommended readings on search:
  - AIMA Ch 3, Ch 5.1-5.3.