CS 561: Uncertainty and Bayes Nets

Uncertainty

- Logical agent performs as per the expectation if
  - Knows enough facts about environment

Unfortunately agents almost never have access to the whole truth of the environment !

Impossible to construct a complete and correct descriptions of how its actions will work

Agents must, therefore, act under uncertainty

Uncertainty

Let action A_t = leave for airport *t* minutes before flight departs
Will A_t get me there on time?

Problems:

1. partial observability (road state, other drivers' plans, etc.)
2. noisy sensors (traffic reports)
3. uncertainty in action outcomes (flat tire, etc.)
4. immense complexity of modeling and predicting traffic

Hence a purely logical approach either

1. risks falsehood: “A_{25} will get me there on time”, or
2. leads to conclusions that are too weak for decision making:

“A_{25} will get me there on time if there's no accident on the bridge and it doesn't rain and my tires remain intact etc etc.”

(A_{1440} might reasonably be said to get me there on time but I'd have to stay overnight in the airport …)

Methods for handling uncertainty

- Default or non-monotonic logic (*consequences may be derived only because of lack of evidence of the contrary*):
  - Assume my car does not have a flat tire
  - Assume A_{25} works unless contradicted by evidence
- Issues: *What assumptions are reasonable? How to handle contradiction?*
- Probability
  - Model agent's degree of belief
  - Given the available evidence, A_{25} will get me there on time with probability 0.04

Probability

- Expresses uncertainty
- Pervasive in many applications of CS
  - Machine learning, Pattern recognition
  - Information Retrieval (e.g., Web)
  - Computer Vision
  - Robotics
- Based on mathematical calculus

- Disclaimer: We only discuss finite distributions

Probability

Probabilistic assertions summarize effects of

- Laziness: failure to enumerate exceptions, qualifications, etc.
- Ignorance: lack of relevant facts, initial conditions, etc.

Subjective probability:

- Probabilities relate propositions to agent's own state of knowledge

  e.g., $P(A_{25} \mid \text{no reported accidents}) = 0.6$

These are not assertions about the world

does not imply that whenever there are no accidents will reach the airport with 0.6

implies: whenever there are no accidents and no other information is available then reach the airport with 0.6

Probabilities of propositions change with new evidence:

e.g., $P(A_{25} \mid \text{no reported accidents, 5 a.m.}) = 0.15$

Making decisions under uncertainty

Suppose I believe the following:

$P(A_{25}$ gets me there on time | …) = 0.04
$P(A_{90}$ gets me there on time | …) = 0.70
$P(A_{120}$ gets me there on time | …) = 0.95
$P(A_{1440}$ gets me there on time | …) = 0.9999

- Which action to choose?

  Depends on my preferences for missing flight vs. time spent waiting, etc.

  - Utility theory is used to represent and infer preferences
  - Decision theory = probability theory + utility theory

Syntax

- Basic element: random variable
- Similar to propositional logic: possible worlds defined by assignment of values to random variables
- Boolean random variables
  e.g., *Cavity* (do I have a cavity?)
- Discrete random variables
  e.g., *Weather* is one of <*sunny,rainy,cloudy,snow*>
- Domain values must be exhaustive and mutually exclusive
- Elementary proposition constructed by assignment of a value to a random variable: e.g., *Weather = sunny*, *Cavity = false* (abbreviated as $\neg$*cavity*)
- Complex propositions formed from elementary propositions and standard logical connectives e.g., *Weather = sunny* $\vee$ *Cavity = false*

Syntax

- Atomic event: A complete specification of the state of the world about which the agent is uncertain

E.g., if the world consists of only two boolean variables *Cavity* and *Toothache*, then there are 4 distinct atomic events:

$Cavity = false \wedge Toothache = false$

$Cavity = false \wedge Toothache = true$

$Cavity = true \wedge Toothache = false$

$Cavity = true \wedge Toothache = true$

Properties of atomic events

- Atomic events are mutually exclusive

  E.g. *Cavity* $\wedge$ *Toothache and Cavity* $\wedge \neg$ *Toothache*
- Set of atomic events is exhaustive-*disjunction of all atomic events is logically equivalent to true*
- Atomic event entails the truth or falsehood of every proposition

  e.g. *Cavity* $\wedge \neg$ *Toothache* entails the truth of *cavity*

  and falsehood of Toothache
- *Logically equivalent to the disjunction of all atomic events that entail the truth of proposition*

  e.g. cavity= (Cavity $\wedge$ *Toothache) V (Cavity* $\wedge \neg$ *Toothache)*

Probability

- Probability of a fair coin

$$P(\mathrm{COIN} = \mathrm{tail}) = \frac{1}{2}$$

$$P(\mathrm{tail}) = \frac{1}{2}$$

Probability

- Probability of cancer

$$P(\text{has cancer}) = 0.02$$

$$\text{Þ} \quad P(\text{Ø}\,\text{has cancer}) = 0.98$$

Joint Probability

- Multiple events: cancer, test result

$$P(\text{has cancer, test positive})$$

Has cancer?	Test positive?	P(C,TP)
yes	yes	0.018
yes	no	0.002
no	yes	0.196
no	no	0.784

Joint Probability

- The problem with joint distributions

It takes 2^D-1 numbers to specify them!

Conditional Probability

- Describes the cancer test:

$$P(\text{test positive} \mid \text{has cancer}) = 0.9$$
$$P(\text{test positive} \mid \neg\text{has cancer}) = 0.2$$

- Put this together with: Prior probability

$$P(\text{has cancer}) = 0.02$$

P(A, B)= P(A|B) * P(B)=P(B|A) * P(A)

Conditional Probability

- We have:

$$P(\mathrm{C}) = 0.02 \qquad P(\neg\mathrm{C}) = 0.98$$

$$P(\mathrm{TP} \mid \mathrm{C}) = 0.9 \qquad P(\neg\mathrm{TP} \mid \mathrm{C}) = 0.1$$

$$P(\mathrm{TP} \mid \neg\mathrm{C}) = 0.2 \qquad P(\neg\mathrm{TP} \mid \neg\mathrm{C}) = 0.8$$

- We can now calculate joint probabilities

Has cancer?	Test positive?	P(TP, C)
yes	yes	0.018
yes	no	0.002
no	yes	0.196
no	no	0.784

Conditional Probability

- "Diagnostic" question: How likely do is cancer given a positive test?

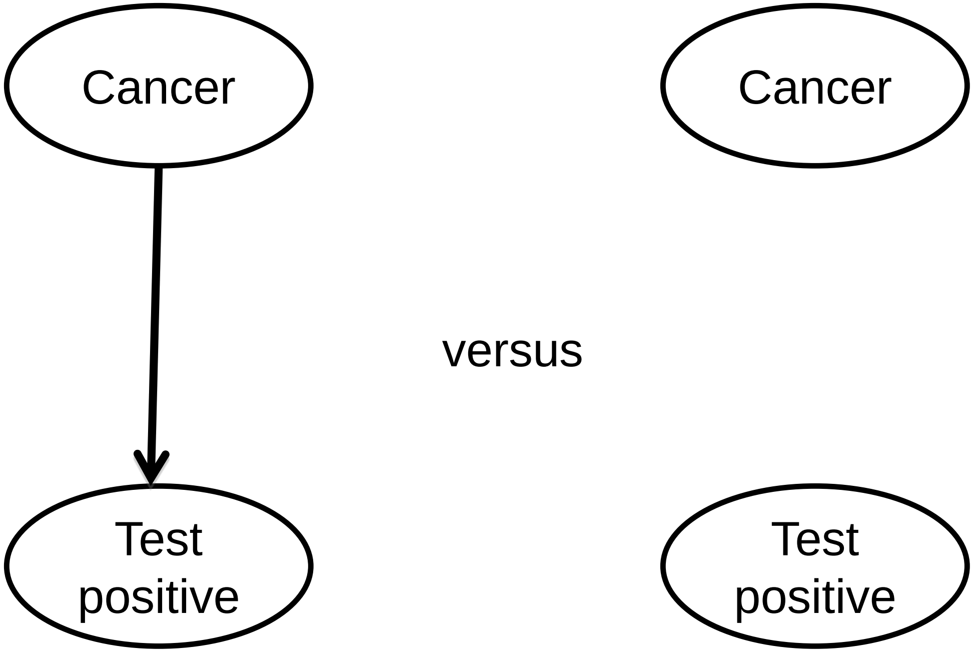
$$P(\text{has cancer} \mid \text{test positive}) = ?$$

Has cancer?	Test positive?	P(TP, C)
yes	yes	0.018
yes	no	0.002
no	yes	0.196
no	no	0.784

$$P(\mathrm{C} \mid \mathrm{TP}) = P(\mathrm{C}, \mathrm{TP}) / P(\mathrm{TP}) = 0.018 / 0.214 = 0.084$$

Bayes Network

- We just encountered our first Bayes network:



Conditional Probability

- Chain rule is derived by successive application of product rule:

$$
\begin{aligned}
\mathbf{P}(X_1, \ldots, X_n) &= \mathbf{P}(X_1,\ldots,X_{n-1})\ \mathbf{P}(X_n \mid X_1,\ldots,X_{n-1}) \\
&= \mathbf{P}(X_1,\ldots,X_{n-2})\ \mathbf{P}(X_{n-1} \mid X_1,\ldots,X_{n-2})\ \mathbf{P}(X_n \mid X_1,\ldots,X_{n-1}) \\
&= \ldots \\
&= \pi_{i=1}\hat{}n\ \mathbf{P}(X_i \mid X_1, \ldots, X_{i-1})
\end{aligned}
$$

Inference by enumeration

- Start with the joint probability distribution:

	toothache		$\neg$ *toothache*	
	catch	$\neg$ *catch*	*catch*	$\neg$ *catch*
cavity	.108	.012	.072	.008
$\neg$ *cavity*	.016	.064	.144	.576

- For any proposition φ, sum the atomic events where it is true: $P(\varphi) = \Sigma_{\omega:\omega \models \varphi} P(\omega)$

Inference by enumeration

- Start with the joint probability distribution:

	toothache		$\neg$ *toothache*	
	catch	$\neg$ *catch*	*catch*	$\neg$ *catch*
cavity	.108	.012	.072	.008
$\neg$ *cavity*	.016	.064	.144	.576

- For any proposition φ, sum the atomic events where it is true: $P(\varphi) = \Sigma_{\omega:\omega \models \varphi} P(\omega)$

$P(toothache) = 0.108 + 0.012 + 0.016 + 0.064 = 0.2$

Inference by enumeration

- Start with the joint probability distribution:

	toothache		*¬ toothache*	
	catch	*¬ catch*	*catch*	*¬ catch*
cavity	.108	.012	.072	.008
¬ cavity	.016	.064	.144	.576

- For any proposition φ, sum the atomic events where it is true: $P(\varphi) = \Sigma_{\omega:\omega \models \varphi} P(\omega)$

P(*toothache*) = 0.108 + 0.012 + 0.016 + 0.064 = 0.2

Inference by enumeration

- Start with the joint probability distribution:

	toothache		¬ *toothache*	
	catch	¬ *catch*	*catch*	¬ *catch*
cavity	.108	.012	.072	.008
¬ *cavity*	.016	.064	.144	.576

- Can also compute conditional probabilities:

$$P(\neg cavity \mid toothache) = \frac{P(\neg cavity \wedge toothache)}{P(toothache)}$$

$$= \frac{0.016+0.064}{0.108 + 0.012 + 0.016 + 0.064}$$

$$= 0.4$$

Normalization

	toothache		*¬ toothache*	
	catch	*¬ catch*	*catch*	*¬ catch*
cavity	.108	.012	.072	.008
¬ cavity	.016	.064	.144	.576

- Denominator can be viewed as a normalization constant α

$\mathbf{P}(Cavity \mid toothache) = \alpha,\ \mathbf{P}(Cavity, toothache)$

$= \alpha,\ [\mathbf{P}(Cavity, toothache, catch) + \mathbf{P}(Cavity, toothache, \neg\, catch)]$

$= \alpha,\ [\langle 0.108, 0.016\rangle + \langle 0.012, 0.064\rangle]$

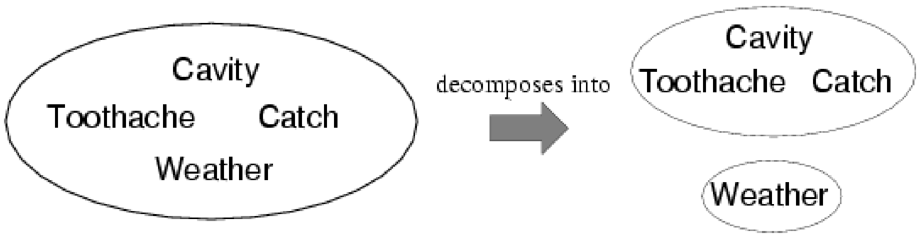
$= \alpha,\ \langle 0.12, 0.08\rangle = \langle 0.6, 0.4\rangle$

$\alpha = P(\text{toothache})$

General idea: compute distribution on query variable by fixing evidence variables and summing over hidden variables

Independence

- *A* and *B* are independent iff
  $\mathbf{P}(A|B) = \mathbf{P}(A)$ or $\mathbf{P}(B|A) = \mathbf{P}(B)$ or $\mathbf{P}(A, B) = \mathbf{P}(A)\,\mathbf{P}(B)$



$$\mathbf{P}(Toothache, Catch, Cavity, Weather)$$
$$= \mathbf{P}(Toothache, Catch, Cavity)\,\mathbf{P}(Weather)$$

- Absolute independence powerful but rare

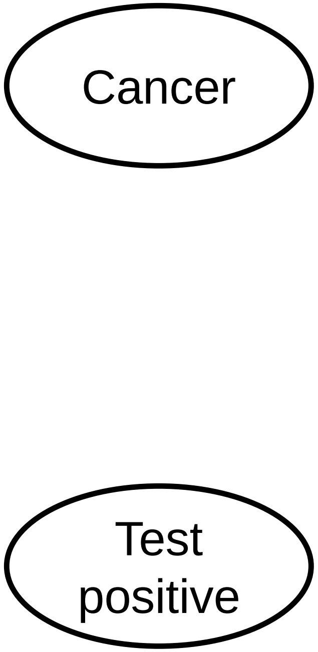
- Dentistry is a large field with hundreds of variables, none of which are independent. What to do?

Independence

- Independence

$$P(\mathrm{C}, \mathrm{TP}) = P(\mathrm{C}) \times P(\mathrm{TP})$$

- What does this mean for our test?
  - Don’t take it!



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 of two simpler distributions
  - This implies:

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

  - We write: $X \perp\!\!\!\perp Y$
- Independence is a simplifying *modeling assumption*
  - *Empirical* joint distributions: at best "close" to independent

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

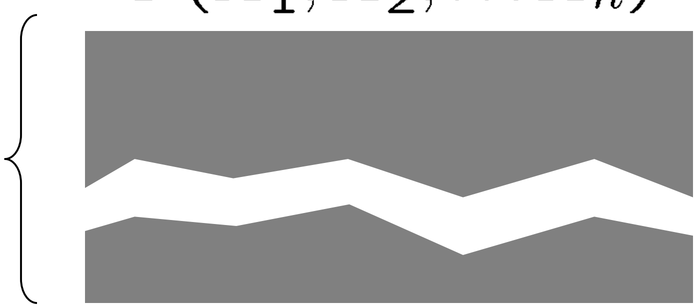
$\ldots$

$P(X_n)$

h	0.5
t	0.5

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

2^n



Example: Independence?

$P(T)$

T	P
warm	0.5
cold	0.5

$P_2(T, W)$

T	W	P
warm	sun	0.3
warm	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 Toothache, a dental probe might be more likely to catch
- But: if I have a cavity, the probability that the probe catches 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, $\neg$cavity) = P(+catch| $\neg$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
- We write: $X \perp\!\!\!\perp Y | Z$

Bayes Network Representation

P(cavity)

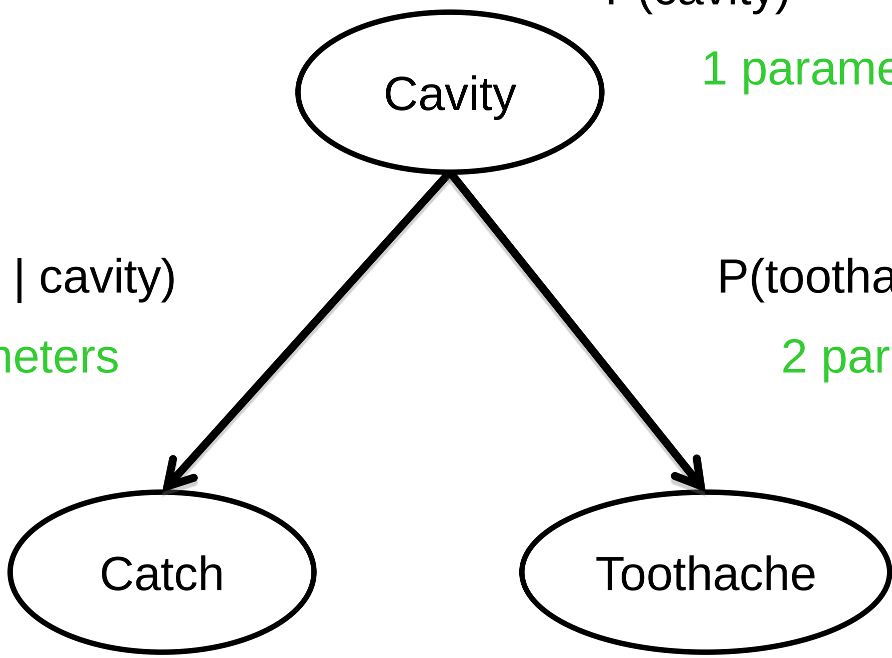
1 parameter

P(catch | cavity)

2 parameters

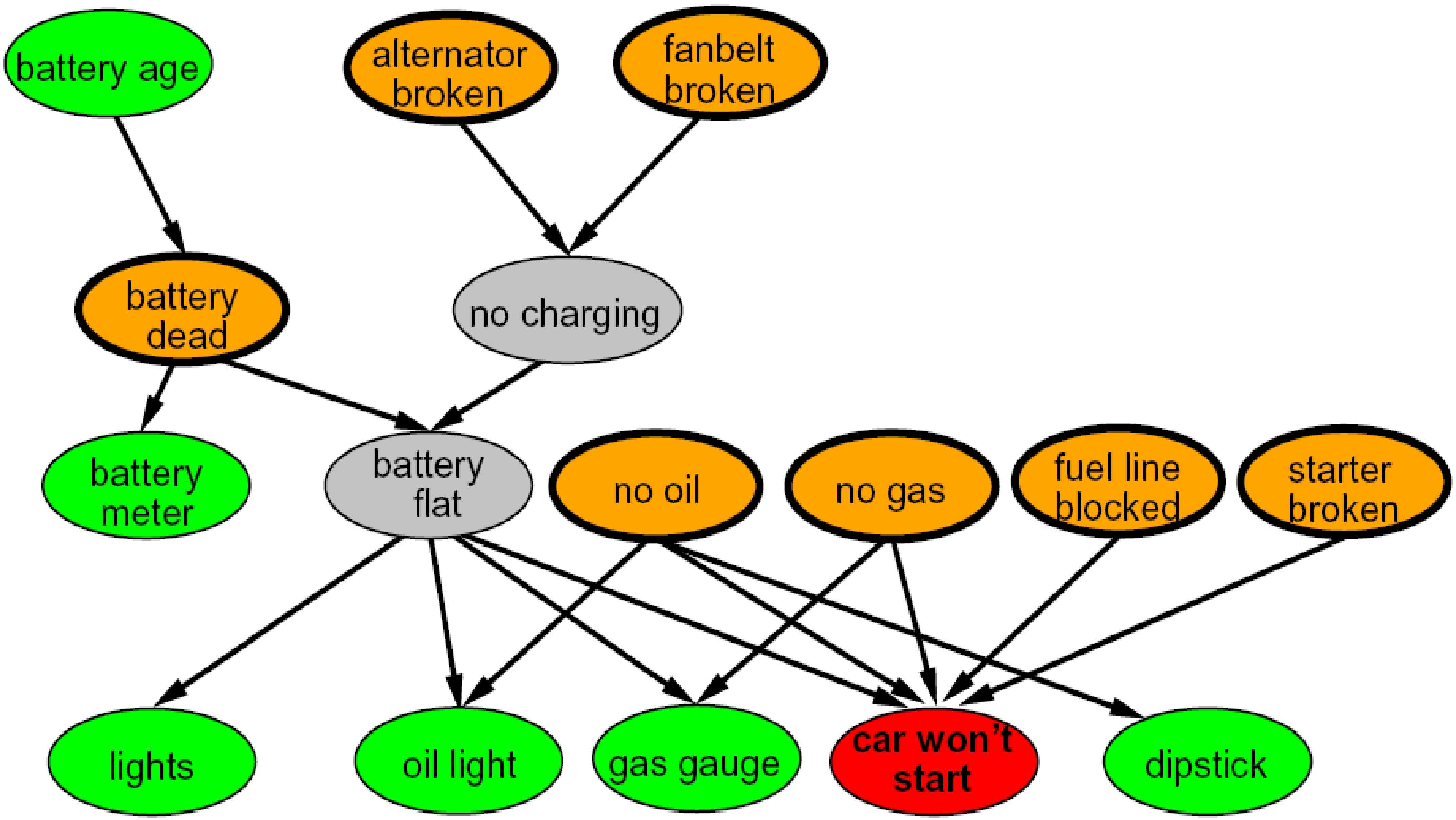
P(toothache | cavity)

2 parameters



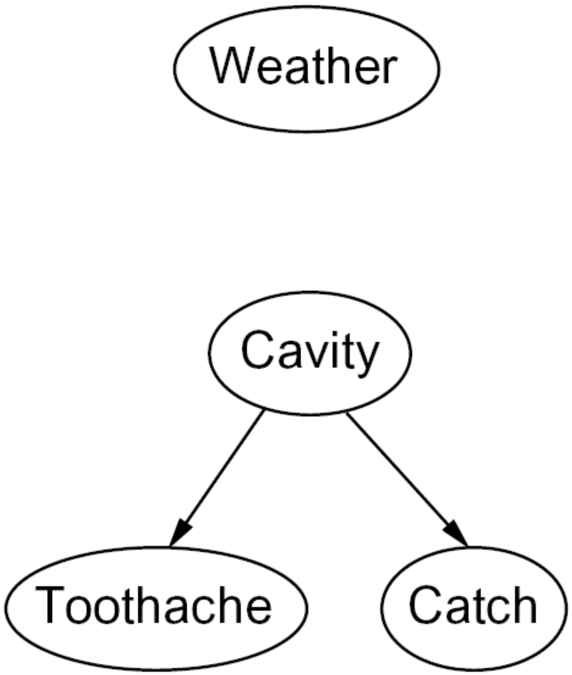
Versus: 2^3-1 = 7 parameters

Example Bayes Network: Car



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 (they may not!)



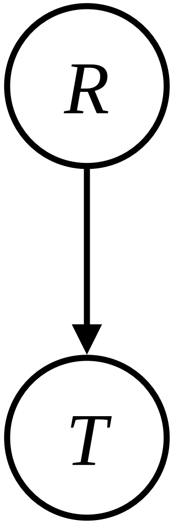
Example: Coin Flips

- N independent coin flips

- No interactions between variables: absolute independence

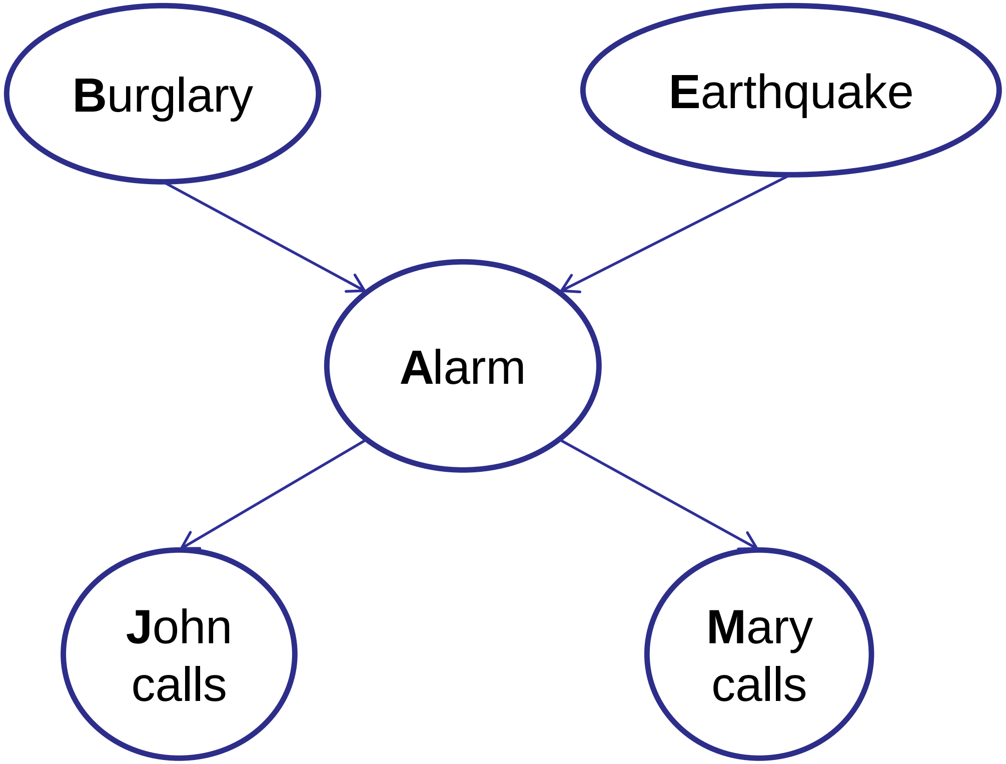
Example: Traffic

- 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: Alarm Network

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



Problem Description

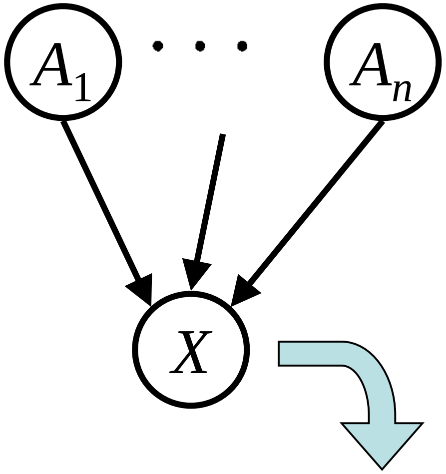
- New burglar alarm installed
- Detects burglary fairly but also responds on occasion to minor earthquake
- John and Mary- two neighbors
  - Calls when they hear the alarm
  - Sometimes confuses the telephone ringing with the alarm and then calls
  - John likes loud music and often misses the alarm together
- Given the evidence of who has or has not called, we would like to estimate the probability of a burglary

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

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

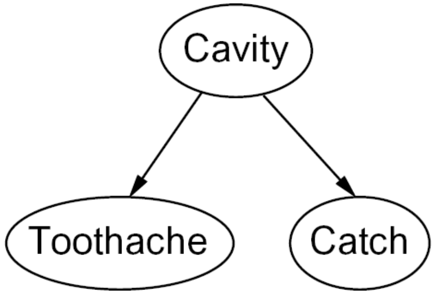
  - CPT: conditional probability table
  - Description of a noisy "causal" process
    - Uncertain relationships



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

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:

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

  - Example:

$$P(+\mathit{cavity},\ +\mathit{catch},\ \neg \mathit{toothache})$$

- This lets us reconstruct any entry of the full joint
- Not every BN can represent every joint distribution
  - The topology enforces certain conditional independencies

Example: Coin Flips

$P(X_1)$

h	0.5
t	0.5

$P(X_2)$

h	0.5
t	0.5

$\cdots$

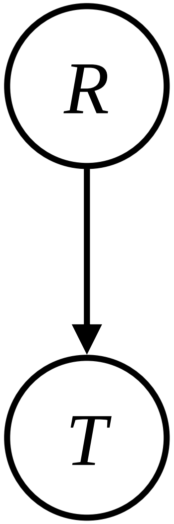
$P(X_n)$

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

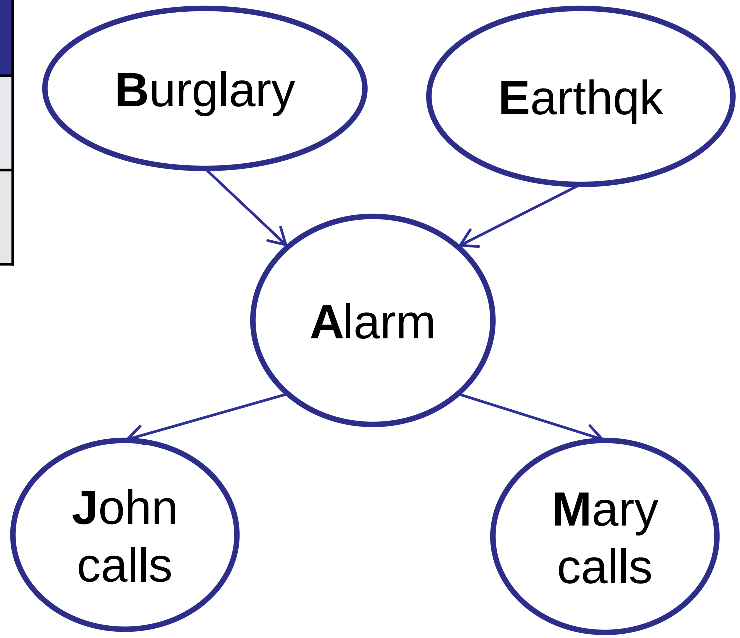
¬r →

+t	1/2
¬t	1/2

$$P(+r, \neg t) =$$

Example: Alarm Network

B	P(B)
+b	0.001
¬b	0.999



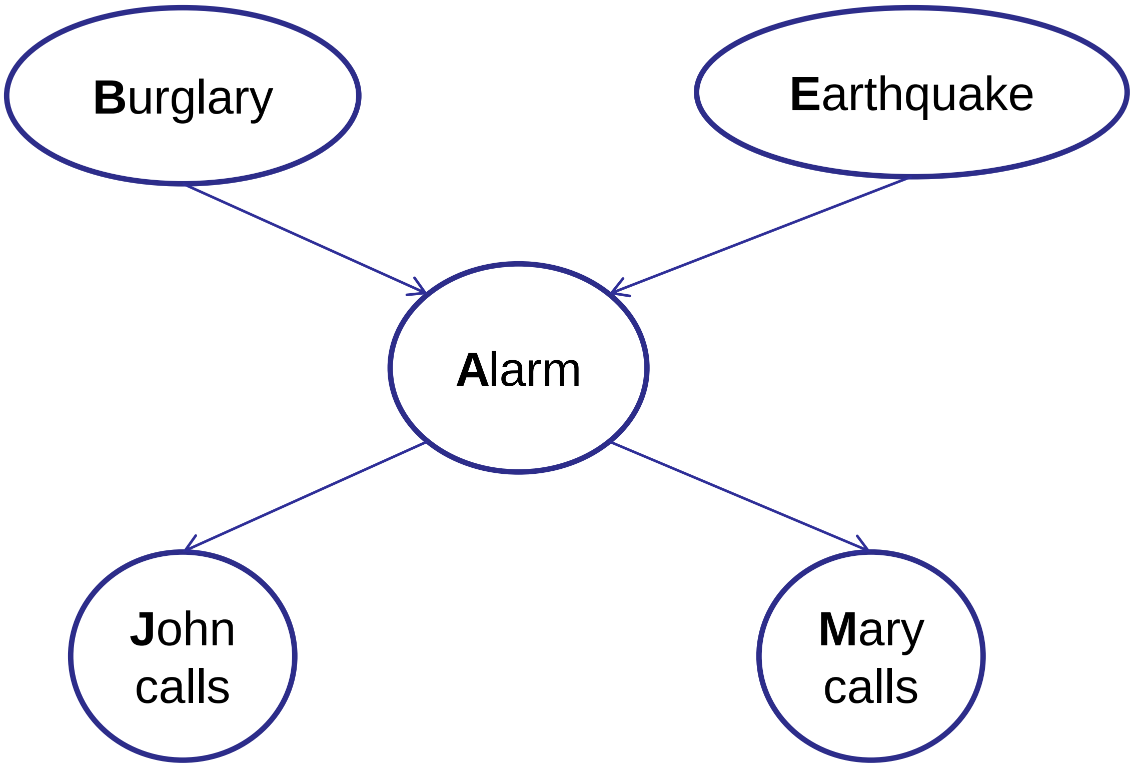
E	P(E)
+e	0.002
¬e	0.998

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

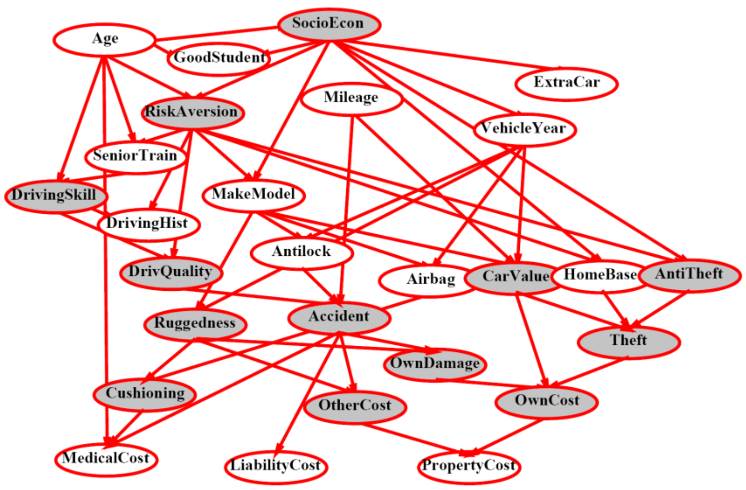
Example: Alarm Network



$$\prod_i P(X_i|\text{Parents}(X_i)) = P(B) \cdot P(E) \cdot P(A|B,E) \cdot P(J|A) \cdot P(M|A)$$

Bayes' Nets

- A Bayes' net is an efficient encoding of a probabilistic model of a domain

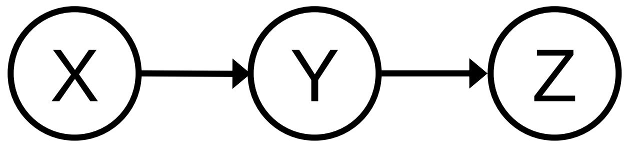


- Questions we can ask:
  - Inference: given a fixed BN, what is P(X | e)?
  - Representation: given a BN graph, what kinds of distributions can it encode?
  - Modeling: what BN is most appropriate for a given domain?

- Find Conditional (In)Dependencies
  - Concept of “d-separation”

Causal Chains

- This configuration is a "causal chain"



X: Low pressure

Y: Rain

Z: Traffic

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

  - Is X independent of Z given Y?

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

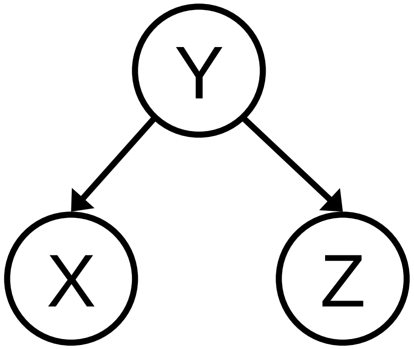
$$= P(z|y)$$

Yes!

  - Evidence along the chain "blocks" the influence

Common Cause

- Another basic configuration: two effects of the same cause
  - Are X and Z independent?
  - Are 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!

Y: Alarm

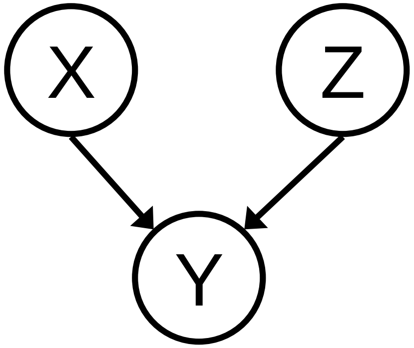
X: John calls

Z: Mary calls

  - Observing the cause blocks influence between effects

Common Effect

- Last configuration: two causes of one effect (v-structures)
  - Are X and Z 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 Z independent given Y?
    - 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



X: Raining

Z: Ballgame

Y: Traffic

The General Case

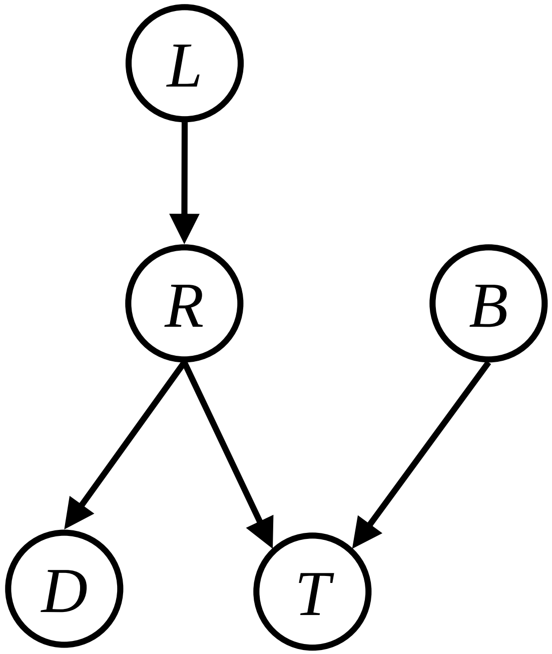
- Any complex example can be analyzed using these three canonical cases
- General question: in a given BN, are two variables independent (given evidence)?
- Solution: analyze the graph

Reachability

- Recipe: shade evidence nodes

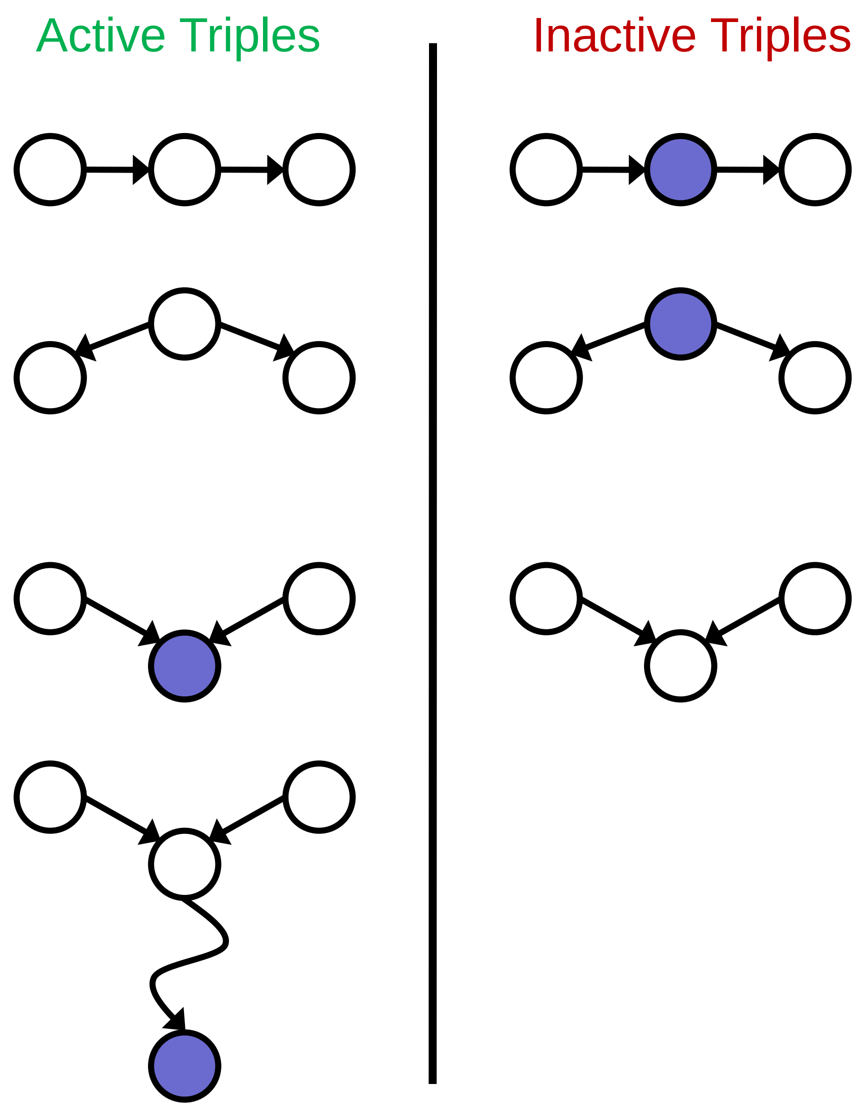
- Attempt 1: Remove shaded nodes. If two nodes are still connected by an undirected path, they are not conditionally independent

- 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"



Reachability (D-Separation)

- Question: Are X and Y conditionally independent given evidence vars {Z}?
  - Yes, if X and Y "separated" by Z
  - Look for active paths from X to Y
  - No active paths = independence!
- A path is active if each triple is active:
  - Causal chain $A \to B \to C$ where B is unobserved (either direction)
  - Common cause $A \leftarrow B \to C$ where B is unobserved
  - Common effect (aka v-structure) $A \to B \leftarrow C$ where B *or one of its descendents* is observed
- All it takes to block a path is a single inactive segment

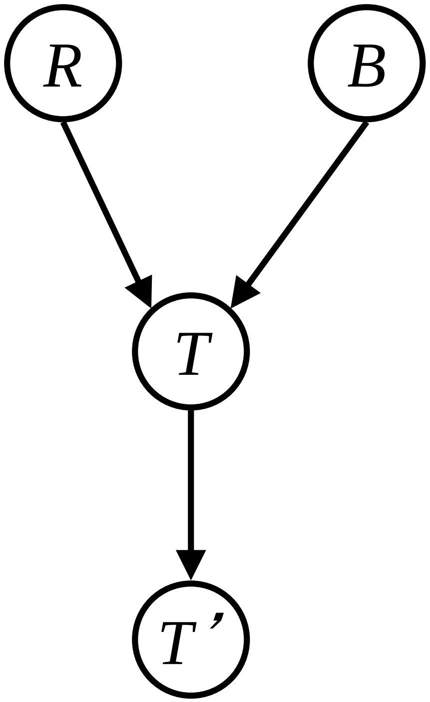


Example

$R \perp\!\!\!\perp B$ *Yes*

$R \perp\!\!\!\perp B | T$

$R \perp\!\!\!\perp B | T'$



Example

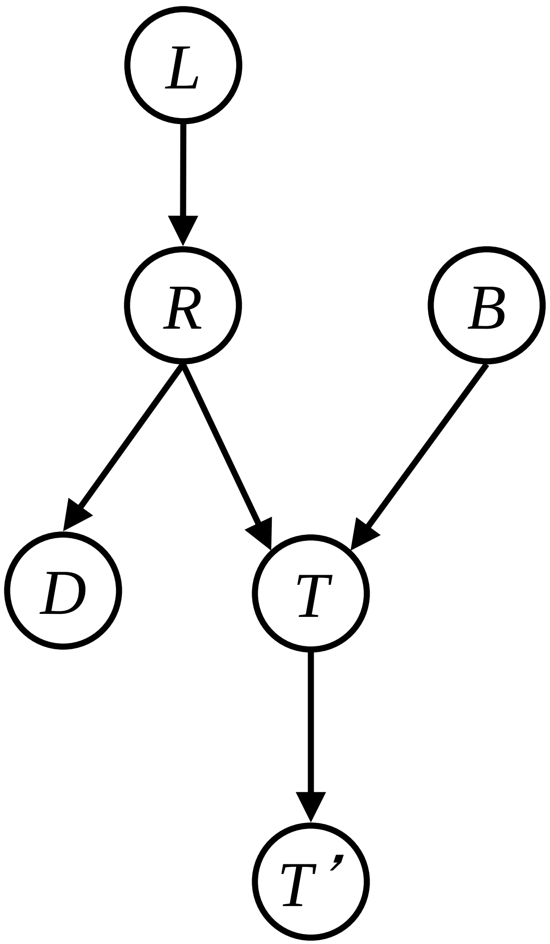
$L \perp\!\!\!\perp T' | T$ Yes

$L \perp\!\!\!\perp B$ Yes

$L \perp\!\!\!\perp B | T$

$L \perp\!\!\!\perp B | T'$

$L \perp\!\!\!\perp B | T, R$ Yes



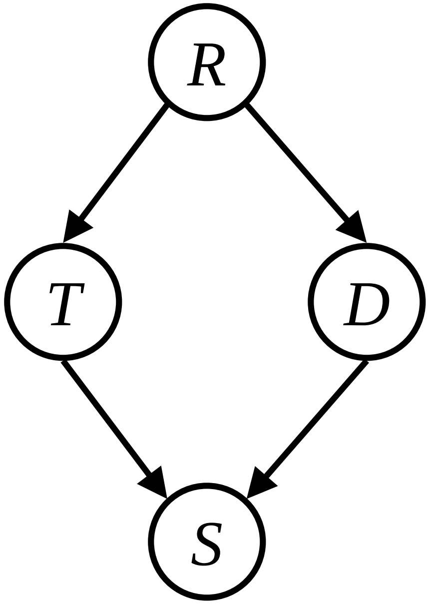
Example

- Variables:
  - R: Raining
  - T: Traffic
  - D: Roof drips
  - S: I'm sad
- Questions:

$$T \perp\!\!\!\perp D$$

$$T \perp\!\!\!\perp D | R$$ *Yes*

$$T \perp\!\!\!\perp D | R, S$$



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
  - End up with arrows that reflect correlation, not causation

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