FIRST ORDER LOGIC AND PROBABILISTIC INFERENCING

Resolution

- Recall from Propositional Logic
  - $$\frac{(\alpha \vee \beta),\ (\neg\beta \vee \gamma)}{(\alpha \vee \gamma)}$$
  - Resolution rule is both sound and complete
- Idea is to prove something using **refutation** (proof by contradiction).
  - We want to prove something (clause/sentence).
  - We add its negation into the KB.
  - We attempt to reach a contradiction as a result of the addition of our negated goal.
  - If contradiction reached, our goal must be true.
- Note: same idea as in propositional logic, but need to handle **quantifiers** and **variables**.

Resolution

Resolution is step-by-step:

1. Convert problem into first order logic expressions.
2. Convert logic expressions into clause form.*
3. Add negation of what we intend to prove (in clause form) to our logic expressions.
4. Use resolution rule to produce new clauses that follow from what we know.
5. Produce a contradiction that proves our goal.

- **Any variable substitutions used during the above steps are those assignments for which the opposite of the negated goal is true.**
  - **Has to do with Unification, more later.**

Conversion to clause form

- Recall resolution for clauses α _ β , : β _ γ resolves to α _ γ.

- So resolution works with pairs (or groups) of disjuncts to produce new disjuncts.

- Therefore, need to convert first order logic into disjuncts (Conjunctive Normal Form, or clause form).

    - Individual clauses *conjuncted* together, each clause made up of *disjunctions* of literals.

Rule Name	Premises	Derived Conclusion
Modus Ponens	$A, A \Rightarrow B$	B
And Introduction	A, B	$A \wedge B$
And Elimination	$A \wedge B$	A
Double Negation	$\neg\neg A$	A
Unit Resolution	$A \vee B, \neg B$	A
Resolution	$A \vee B, \neg B \vee C$	$A \vee C$

Conversion procedure

1. Eliminate all) using the fact that A) B ´ : **A _ B**.

2. Reduce scope of negation to the predicate level using:

    a) **: (: A) ´ A**
    b) **: (9 X), A(X) ´ (8 X), : A(X)**
    c) **: (8 X), A(X) ´ (9 X), : A(X)**
    d) **: (A ^ B) ´ : A _ : B**
    e) **: (A _ B) ´ : A ^ : B**

3. Rename variables so that they are different if bound by different quantifiers.

Conversion procedure

4. Eliminate existential quantifiers. This is known as **Skolemization**.

   a) E.g., **9 X, DOG(X)** may be replaced by **DOG(fido)** where **fido** is a **Skolem constant**.

   b) E.g., 8 **X (9 Y MOTHER(X,Y))** must be replaced by 8 **X MOTHER(X,m(X))** where **m(X)** is a **Skolem function of X**.

5. Drop all 8 universal quantifiers. Assume all variables to be universally quantified.

6. Convert each expression to a conjunction of disjuncts using ***A* _ (B ^ *C*)** ´ **(*A* _ B) ^ (*A* _ *C*)**.

7. Split conjuncts into separate clauses and rename variables in different clauses.

Example of conversion

- Convert: "Everyone who loves all animals is loved by someone"

  8 X [8 Y Animal(Y)) Loves(X,Y)]) [9 Y Loves(Y,X)]

1. Remove) from expression:

   8 X [:8 Y :Animal(Y) _ Loves(X,Y)] _ [9 Y Loves(Y,X)]

2. Reduce scope of negation:

   8 X [9 Y Animal(Y) ^ :Loves(X,Y)] _ [9 Y Loves(Y,X)]

3. Rename variables:

   8 X [9 Y Animal(Y) ^ :Loves(X,Y)] _ [9 Z Loves(Z,X)]

4. Eliminate **9** using Skolemization:

   8 X [Animal(F(X)) ^ :Loves(X, F(X))] _ [Loves(G(X), X)]

Example of conversion

5. Drop universal quantifiers:
   [Animal(F(X)) ^ :Loves(X, F(X))] _ [Loves(G(X), X)]

6. Convert to conjuction of disjuncts:
   [Animal(F(X)) _ Loves(G(X), X)] ^ [:Loves(X, F(X)) _ Loves(G(X), X)]

7. Separate into clauses:
   [Animal(F(X)) _ Loves(G(X), X)]
   [:Loves(X, F(X)) _ Loves(G(X), X)]

8. Rename variables (again) in different clauses:
   [Animal(F(X)) _ Loves(G(X) ,X)]
   [:Loves(W, F(W)) _ Loves(G(W), W)]

Example – Resolution Refutation Proof

KB:

$\forall x\ P(x) \Rightarrow Q(x)$

$\forall x\ \neg P(x) \Rightarrow R(x)$

$\forall x\ Q(x) \Rightarrow S(x)$

$\forall x\ R(x) \Rightarrow S(x)$

- Prove:

  $S(A)$

- Steps:

  1) Convert KB to CNF
  2) Negate wff we are trying to prove: $\neg S(A)$
  3) Convert negated wff to CNF: $\neg S(A)$

     Add this new clause to KB
  4) Resolve until you get an empty clause

KB in clausal (CNF) form:

$\neg P(x) \vee Q(x)$

$P(y) \vee R(y)$

$\neg Q(z) \vee S(z)$

$\neg R(t) \vee S(t)$

$\neg S(A)$

Resolution steps:

$\neg P(x) \vee Q(x) \qquad P(y) \vee R(y)$

$\neg Q(z) \vee S(z) \qquad Q(a) \vee R(a)$

$\neg R(t) \vee S(t) \qquad S(b) \vee R(b)$

$S(c) \qquad \neg S(A)$

[]

What does this mean?

Example of resolution refutation

- Consider the following story:

  "Anyone passing his or her artificial intelligence (ai) exam and winning the lottery is happy. But anyone who studies or is lucky can pass all his exams. Pete did not study but is lucky. Anyone who is lucky wins the lottery. Is Pete happy?".

- Use:
  - Pass(x,y), Win(x,y), Happy(x), Study(x), Lucky(x), ai, Lottery, Bob

Example of resolution refutation

Step 1: Change sentences to first order logic:

1. "Anyone passing his or her artificial intelligence exam and winning the lottery is happy"
   (8X) (PASS(X,ai) ^ WIN(X,lottery)) HAPPY(X))

2. "Anyone who studies or is lucky can pass all his exams"
   (8 X 8 Y) (STUDIES(X) _ LUCKY(X)) PASS(X,Y))

3. "Pete did not study but is lucky"
   : STUDY(pete) ^ LUCKY(pete)

4. "Anyone who is lucky wins the lottery"
   (8 X) (LUCKY(X)) WINS(X,lottery))

Example of Resolution Refutation

Step 2: Convert all sentences into clause form:

1. (8 X) (PASS(X,ai) ^ WIN(X,lottery)) HAPPY(X)) *gives:*

   : PASS(X,ai) _ : WIN(X,lottery) _ HAPPY(X)

   P= (PASS(X,ai) ^ WIN(X,lottery))

2. (8 X 8 Y) (: STUDIES(X) _ LUCKY(X)) PASS(X,Y)) *gives:*

   : STUDIES(Y) _ PASS(Y,Z)

   : LUCKY(W) _ PASS(W,V)

Recall: $P \Rightarrow Q$ is $(\neg P) \vee Q$

3. :STUDY(pete) ^ LUCKY(pete) *gives:*

   : STUDIES(pete)

   LUCKY(pete)

4. (8 X) (LUCKY(X)) WINS(X,lottery)) *gives:*

   : LUCKY(U) _ WINS(U,lottery)

Example of resolution refutation

Step 3: Add negation (in clause form) of what we want to know:

: HAPPY(pete)

Step 4: Use resolution (and our negated goal) to build a resolution refutation graph to prove a contradiction.

- When building the graph to prove the contradiction, it should be the case that the negative of the goal gets used somewhere in the proof.
- Contradiction occurs when resolve clauses like *A*, : *A*
-
- result = {} (NULL set) since *A* cannot be true and false at the same time!

! PASS(X,ai) OR ! WIN(X,lotter) OR HAPPY(X)

WIN(U,lottery) OR ! LUCKY(U)

Recall: A OR B,:B OR C→A OR C

In this case→--→----→------------→

A=(!PASS(x,ai) OR HAPPY(X))

! PASS(U,ai) OR HAPPY(U) OR ! LUCKY(U)

! HAPPY(pete)

LUCKY(pete)

! PASS(pete,ai) OR ! LUCKY(pete)

Q

:P

P

! PASS(pete,ai)

! LUCKY(V) OR PASS(V,W)

$P \Rightarrow Q$ is $(\neg P) \vee Q$

LUCKY(pete)

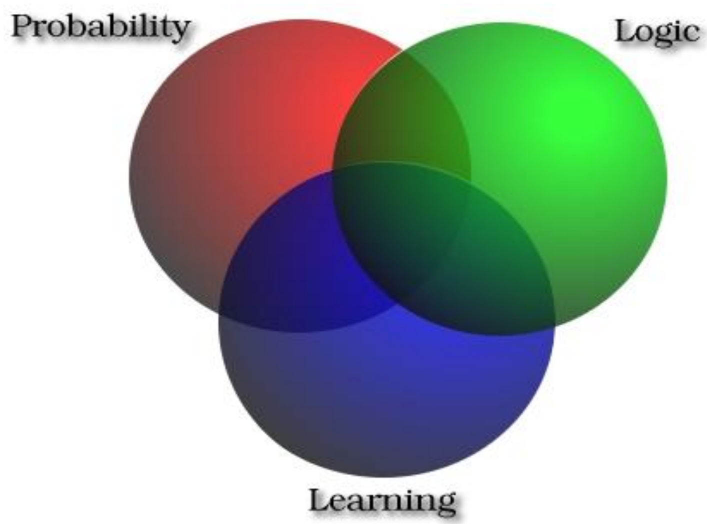
! LUCKY(pete)

{}

Probabilistic Logic Learning

Probabilistic Approach to Uncertainty

- Logic agents almost never have access to the whole truth about their environment.
- There will always be questions to which a categorical answer (i.e., TRUE of FALSE) cannot be found.
- Will try to apply **probability theory** to deal with degree of belief about things.



Bayes Theorem

- We have that $P(A,B) = P(A|B)\,P(B) = P(B|A)\,P(A)$ and therefore we can remove the joint probability to find that:

  $$P(A|B) = P(B|A)\,P(A) / P(B).$$

- Called **Bayes' Theorem** and provides a way to determine a conditional probability without the joint probability of A and B.

Bayes Theorem

- Common to think of Bayes' Theorem in terms of updating our belief about a hypothesis A in the light of new evidence B.
  - Specifically, *our posterior belief P(A|B) is calculated by multiplying our prior belief P(A) by the likelihood P(B|A) that B will occur if A is true.*
- In many situations it is difficult to compute P(A|B) directly, yet we might have information about P(B|A).
  - Bayes' Theorem enables us to compute P(A|B) in terms of P(B|A).
  - This is what makes Bayes' Theorem so powerful.

Example of Bayes Theorem

- Let A represent the event "Person has cancer" and let B represent the event "Person is a smoker“.

- We know the probability of the prior event A is 0.1 based on past data (10% of patients entering the clinic turn out to have cancer). Thus $P(A)=0.1$

- Want to compute the probability of the posterior event $P(A|B)$.

Example of Bayes Theorem

- It is difficult to find this out directly. However, …
  - We are likely to know P(B) by considering the percentage of patients who smoke. Suppose P(B)=0.5.
  - We are also likely to know P(B|A) by checking from our records the proportion of smokers among those diagnosed with cancer. Suppose P(B|A)=0.8.

- We can now use Bayes' Theorem to compute P(A|B) = (0.8   0.1)/0.5 = 0.16

- Thus, in the light of *evidence* that the person is a smoker we revise our prior probability from 0.1 to a posterior probability of 0.16. This is a significance increase, but it is still unlikely that the person has cancer.

Bayes Theorem In a Different Form

- The denominator of Bayes' Theorem P(A|B) = P(B|A) P(A) / P(B) is P(B) and is a normalizing constant which can be computed using marginalization.

$$P(B) = \sum_{1}^{n} P(B|A_i)P(A_i)$$

- Hence we can state Bayes' Theorem in a different form:

$$P(A|B) = \frac{P(B|A)P(A)}{\sum_i P(B|A_i)P(A_i)}$$

Example of Bayes Theorem In a Different Form

- Suppose that we have two bags each containing black and white balls.
- One bag contains three times as many white balls as blacks. The other bag contains three times as many black balls as white.
- Suppose we choose one of these bags at random. For this bag we select five balls at random, replacing each ball after it has been selected. The result is that we find 4 white balls and one black.
- What is the probability that we were using the bag with mainly white balls?

Example of Bayes Theorem In a Different Form

- Let A be the random variable "bag chosen" then A={a1,a2} where a1 represents "bag with mostly white balls" and a2 represents "bag with mostly black balls" . We know that P(a1)=P(a2)=1/2 since we choose the bag at random.

- Let B be the event "4 white balls and one black ball chosen from 5 selections". We have to calculate P(a1|B).

- For the bag with mostly white balls, we find P(B|a1) = 405/1024. For the bag with mostly black balls we find P(B|a2) = 15/1024.

- Hence, P(a1|B) = P(B|a1)P(a1)/(P(B|a1)P(a1)|P(B|a2)P(a2)) = ((405/1024)/(405/1024)+(15/1024)) = 0.964.

Chain Rule

- Recall that $P(A,B) = P(A|B)\,P(B)$.
- Can extend this formula to more variables.
- Example: For 3-variables:

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

- Example: For n-variables:

  $$P(A1,\,A2,\,\ldots,\,An) = P(A1|\,A2,\,\ldots,\,An)\,P(A2|\,A3,\,\ldots,\,An)\,P(An\text{-}1|An)\,P(An)$$

Chain Rule

- In general we refer to this as the **chain rule**.
- This formula provides a means of calculating the full joint probability distribution.
- This formula is especially significant for **Bayesian Belief Nets** where/when many of the variables are conditionally independent (and the formula can be simplified.

Independence and Conditional Independence

- Two events are independent if $P(A \wedge B) = P(A) P(B)$.
- If both events have a positive probability, then the statement of independence of events is equivalent to a statement of conditional independence:

$$P(A|B) = P(A) \text{ if and only if } P(B|A) = P(B) \text{ if and only if } P(A \wedge B) = P(A) P(B)$$

- Can think of independence in the following way: knowledge that one event has occurred does not change the probability assigned to the other event.

Bayesian Networks

Bayesian networks

- A simple, graphical notation for conditional independence assertions and hence for compact specification of full joint distributions

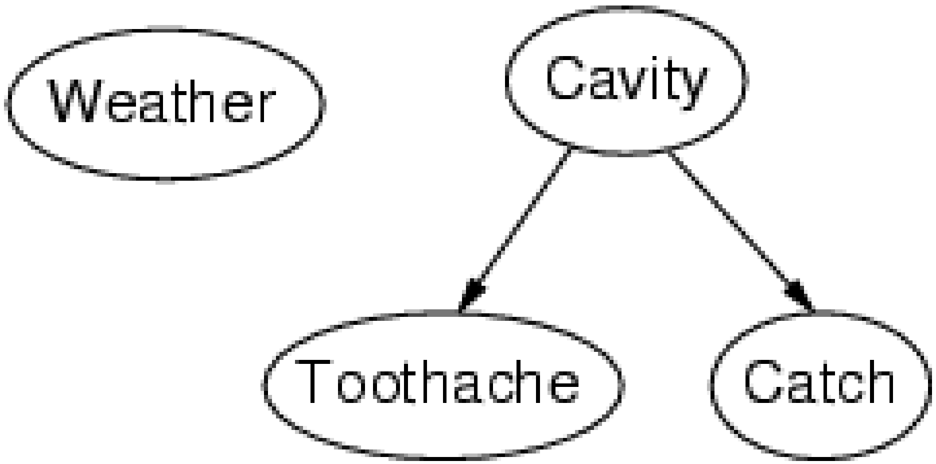
- Syntax:
  - a set of nodes, one per variable
  - a directed, acyclic graph (link ≈ "directly influences")
  - a conditional distribution for each node given its parents:

$$\mathbf{P}\,(X_i \mid Parents\,(X_i))$$

- In the simplest case, conditional distribution represented as a conditional probability table (CPT) giving the distribution over X_i for each combination of parent values

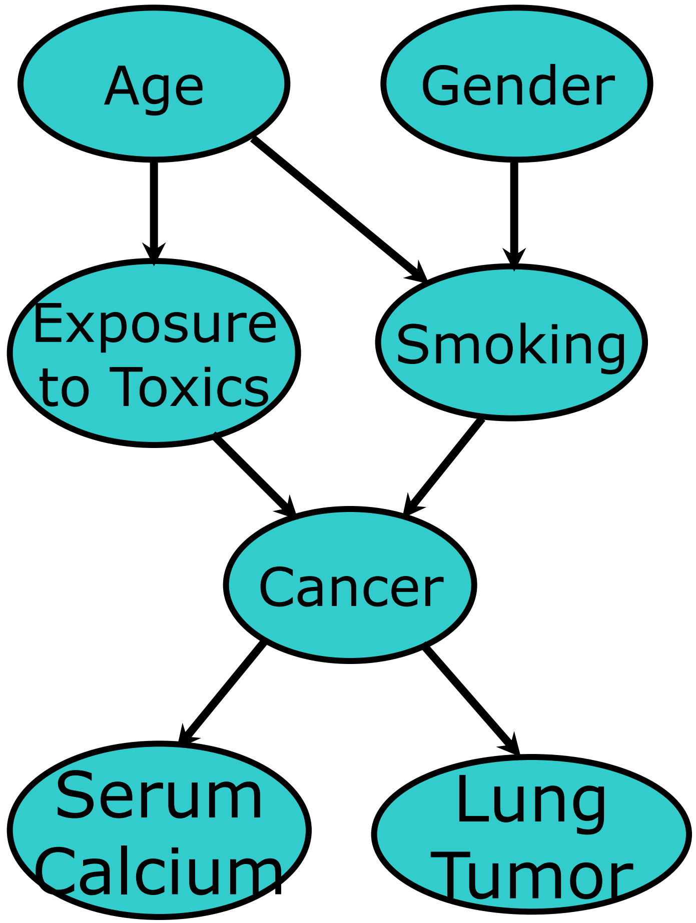
Example

- Topology of network encodes conditional independence assertions:

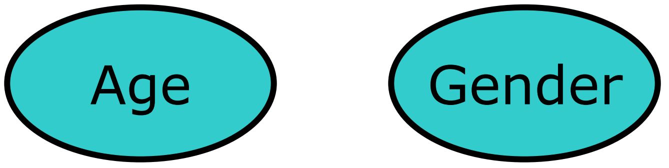


- *Weather* is independent of the other variables
- *Toothache* and *Catch* are conditionally independent given *Cavity*

A Bayesian Network



Independence



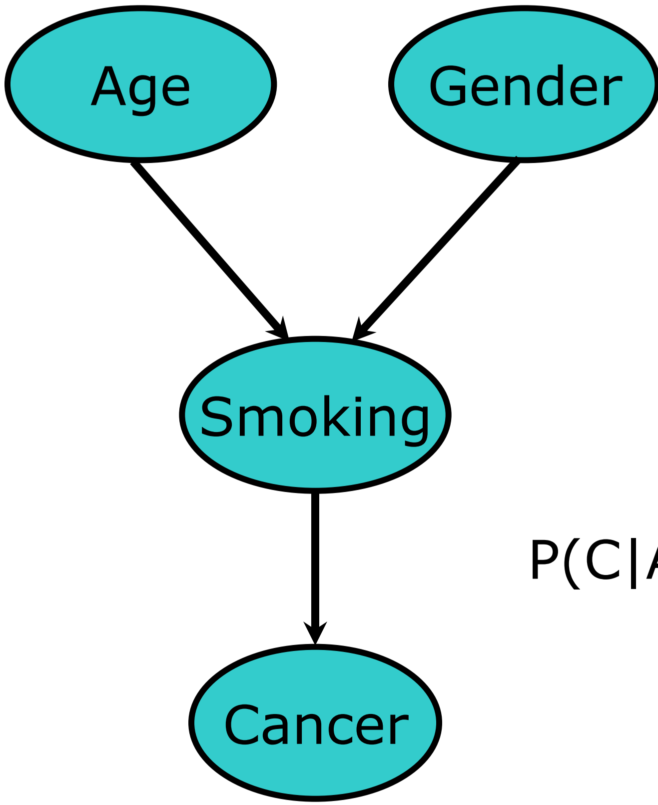
Age and Gender are independent.

$$P(A,G) = P(G)P(A)$$

$$P(A|G) = P(A) \quad A \perp G$$
$$P(G|A) = P(G) \quad G \perp A$$

$$P(A,G) = P(G|A)\,P(A) = P(G)P(A)$$
$$P(A,G) = P(A|G)\,P(G) = P(A)P(G)$$

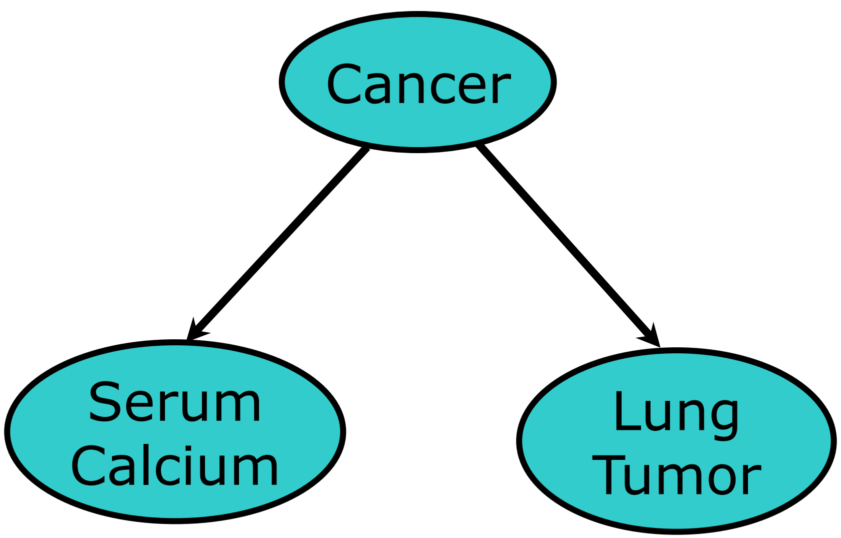
Conditional Independence



Cancer is independent of Age and Gender given Smoking.

$P(C|A,G,S) = P(C|S)$ $\quad C \perp A,G \mid S$

More Conditional Independence: Naïve Bayes

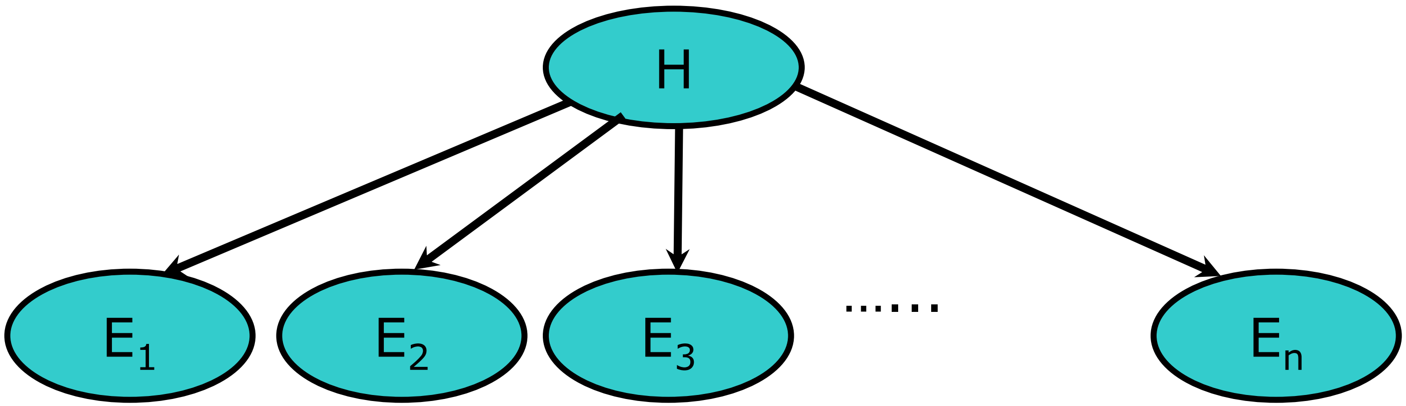


Serum Calcium and Lung Tumor are dependent

Serum Calcium is independent of Lung Tumor, given Cancer

P(L|SC,C) = P(L|C)

Naïve Bayes in general

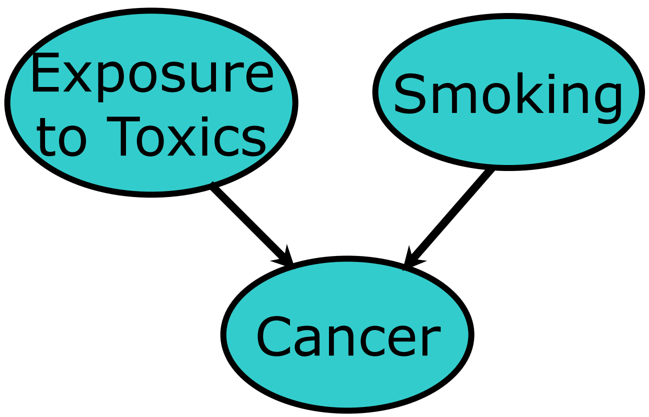


2n + 1 parameters:

$$P(h)$$

$$P(e_i \mid h), P(e_i \mid \bar{h}), \quad i = 1, \dots, n$$

More Conditional Independence: Explaining Away



Exposure to Toxics and Smoking are independent

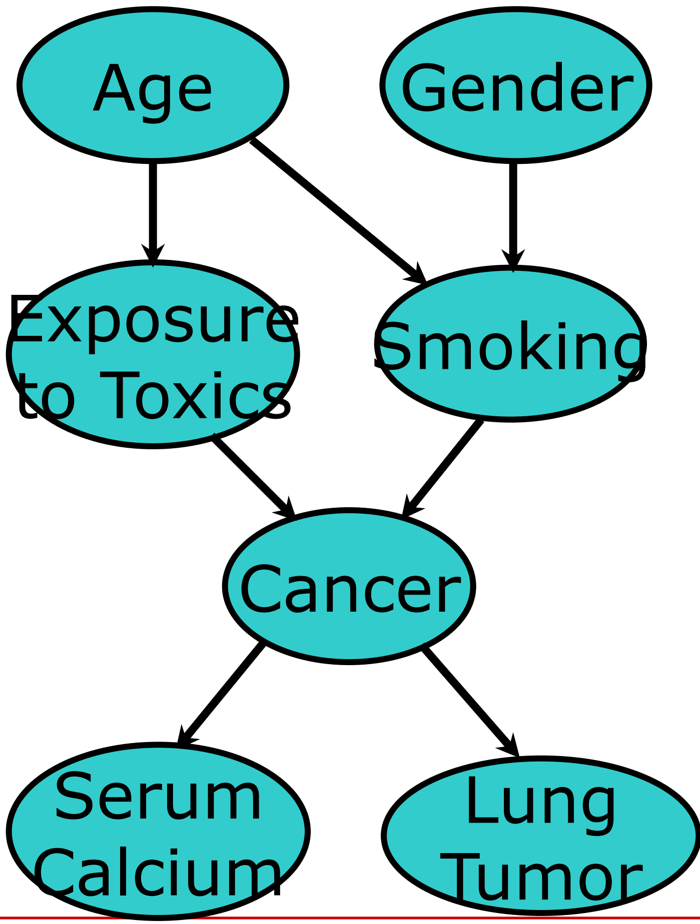
$$E \perp S$$

Exposure to Toxics is **dependent** on Smoking, given Cancer

$$P(E = \text{heavy} \mid C = \text{malignant}) >$$

$$P(E = \text{heavy} \mid C = \text{malignant}, S=\text{heavy})$$

Put it all together



$$P(A, G, E, S, C, L, SC) =$$

$$P(A) \cdot P(G) \cdot$$

$$P(E \mid A) \cdot P(S \mid A, G) \cdot$$

$$P(C \mid E, S) \cdot$$

$$P(SC \mid C) \cdot P(L \mid C)$$

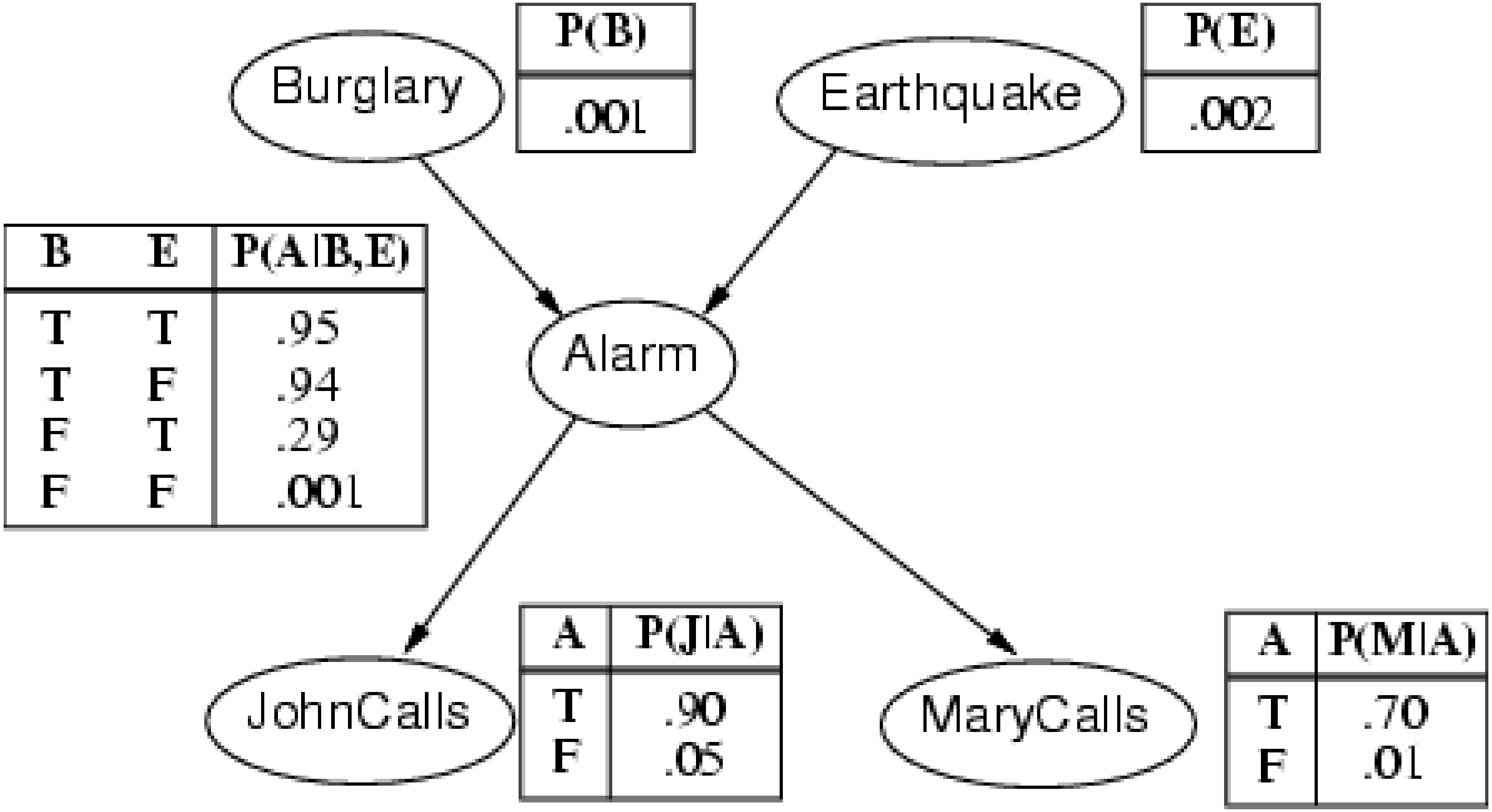
Example

- I'm at work, neighbor John calls to say my alarm is ringing, but neighbor Mary doesn't call. Sometimes it's set off by minor earthquakes. Is there a burglar?

- Variables: *Burglary*, *Earthquake*, *Alarm*, *JohnCalls*, *MaryCalls*

- Network topology reflects "causal" knowledge:
  - A burglar can set the alarm off
  - An earthquake can set the alarm off
  - The alarm can cause Mary to call
  - The alarm can cause John to call

Example contd.



P(B)
.001

P(E)
.002

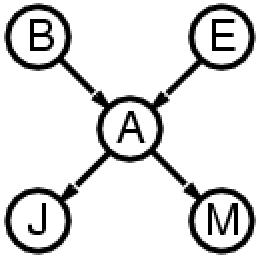
B	E	P(A|B,E)
T	T	.95
T	F	.94
F	T	.29
F	F	.001

A	P(J|A)
T	.90
F	.05

A	P(M|A)
T	.70
F	.01

Compactness

- A CPT for Boolean X_i with k Boolean parents has 2^k rows for the combinations of parent values



- Each row requires one number p for $X_i = true$
  (the number for $X_i = false$ is just $1-p$)

- If each variable has no more than k parents, the complete network requires $O(n \cdot 2^k)$ numbers

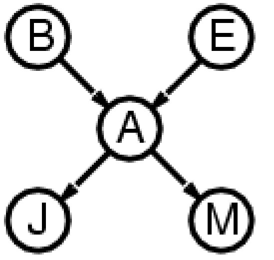
- I.e., grows linearly with n, vs. $O(2^n)$ for the full joint distribution

- For burglary net, $1 + 1 + 4 + 2 + 2 = 10$ numbers (vs. $2^5-1 = 31$)

Semantics

The full joint distribution is defined as the product of the local conditional distributions:

$$\mathbf{P}(X_1, \ldots, X_n) = \pi_{i=1}^{n} \mathbf{P}(X_i \mid Parents(X_i))$$



e.g., $\mathbf{P}(j \wedge m \wedge a \wedge \neg b \wedge \neg e)$

$$= \mathbf{P}(j \mid a)\,\mathbf{P}(m \mid a)\,\mathbf{P}(a \mid \neg b, \neg e)\,\mathbf{P}(\neg b)\,\mathbf{P}(\neg e)$$

Constructing Belief Networks

- Only stated briefly…
- To construct a belief network, you need to consider several things:
  - What are the relevant variables, and what should the domains be? (Defines the nodes).
  - What is the relationship between the variables? (Defines the arcs).
  - How does the value of one variable depend on the variables that locally influence it (i.e., its parents)? (Defines the CPTs).
- Basically, should form an order to adding variables (and therefore edges) to the network. Begin with the root causes, and continue from there.

More on Understanding Conditional Independence

- Can understand dependence in a network by considering how evidence is transmitted in a belief network.

- Can consider several types of connections that exist in a network:

  - **Serial connection** – information entered at the beginning of the connection can be propagated down the connection provided there is no intermediate node has evidence (in which case we stop the propagation).

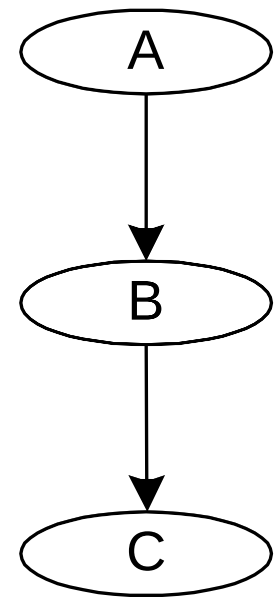
  - Diverging connection – information can be propagated between two child nodes provided the parent node has no evidence.

  - **Converging connection** – information can be propagated between parent nodes when the child node has evidence.

- Rules for these connections are sufficient to describe a complete procedure for determining if two nodes in a belief network are dependent or not. This is formally known as d-separation.

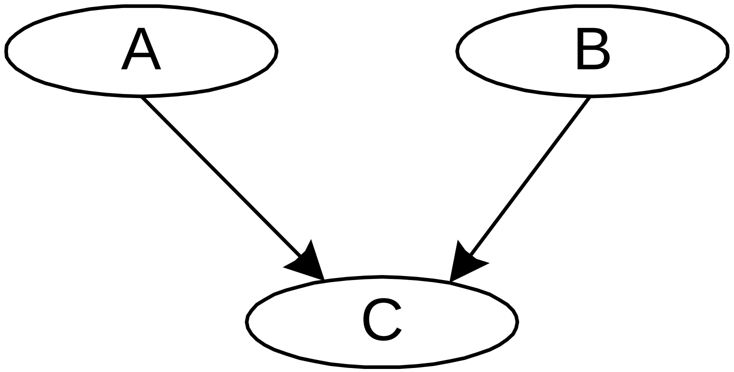
Serial Connection

- In the example below:
  - A and C are dependent.
  - A and C are independent given B (evidence for B).
  - Intuitively, the only way for A to affect C is by affecting B.



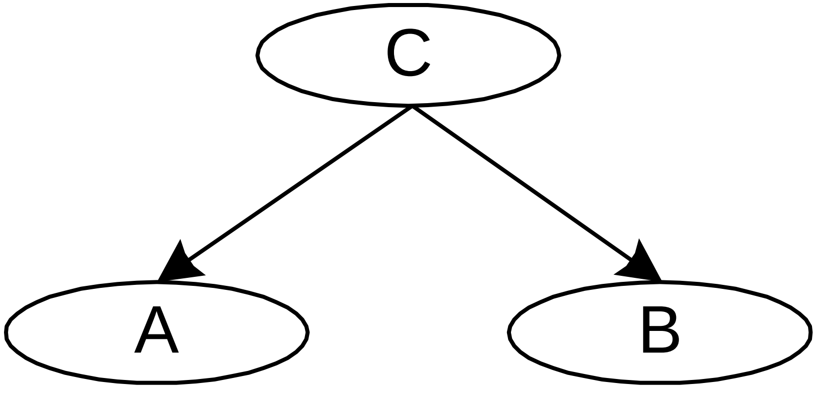
Converging Connection

- In the example below:
  - A and B are independent.
  - A and B are dependent given C.
  - Intuitively, A can be used to explain away B.



Diverging Connection

- In the example below:
  - A and B are dependent.
  - A and B are independent given C.
  - Intuitively, C can explain A and B, meaning one can only affect the other by changing the belief in C.



Constructing Bayesian networks

- 1. Choose an ordering of variables $X_1, \ldots, X_n$
- 2. For $i = 1$ to n
  - add X_i to the network
  -
  - select parents from $X_1, \ldots, X_{i-1}$ such that

$$P(X_i \mid Parents(X_i)) = P(X_i \mid X_1, \ldots X_{i-1})$$

This choice of parents guarantees:

$$P(X_1, \ldots, X_n) = \pi_{i=1}^{n} P(X_i \mid X_1, \ldots, X_{i-1})$$

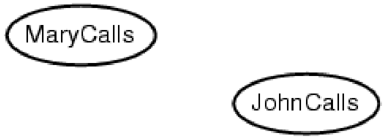
(chain rule)

$$= \pi_{i=1}^{n} P(X_i \mid Parents(X_i))$$

(by construction)

Example

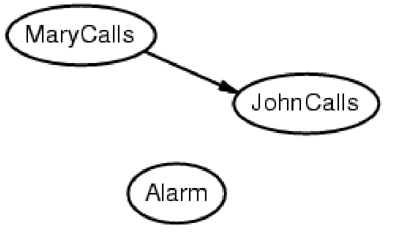
- Suppose we choose the ordering *M, J, A, B, E*
-



$P(J \mid M) = P(J)$?

Example

- Suppose we choose the ordering M, J, A, B, E
-



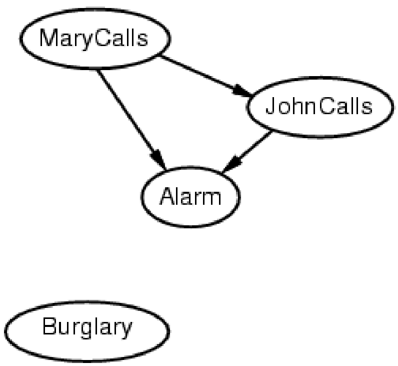
$P(J \mid M) = P(J)$?

No

$P(A \mid J, M) = P(A \mid J)$? $P(A \mid J, M) = P(A)$?

Example

- Suppose we choose the ordering M, J, A, B, E
-



$P(J \mid M) = P(J)$?

No

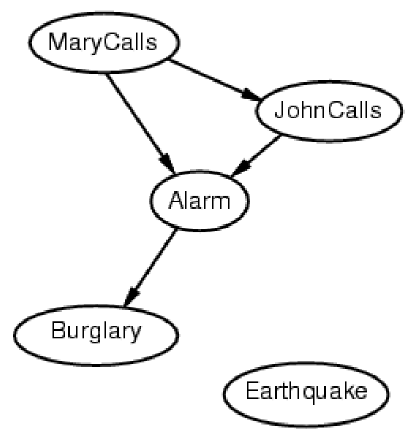
$P(A \mid J, M) = P(A \mid J)$? $P(A \mid J, M) = P(A)$? **No**

$P(B \mid A, J, M) = P(B \mid A)$?

$P(B \mid A, J, M) = P(B)$?

Example

- Suppose we choose the ordering M, J, A, B, E
-



$P(J \mid M) = P(J)$?

No

$P(A \mid J, M) = P(A \mid J)$? $P(A \mid J, M) = P(A)$? **No**

$P(B \mid A, J, M) = P(B \mid A)$? **Yes**

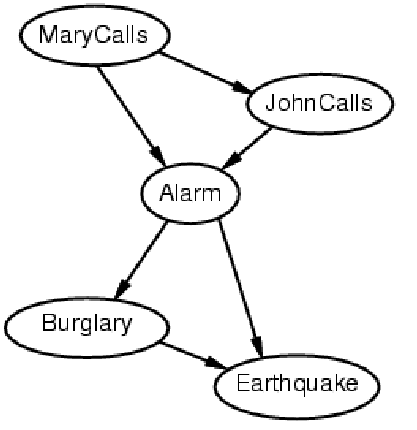
$P(B \mid A, J, M) = P(B)$? **No**

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

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

Example

- Suppose we choose the ordering M, J, A, B, E
-



P(J | M) = P(J)?

No

P(A | J, M) = P(A | J)? P(A | J, M) = P(A)? **No**

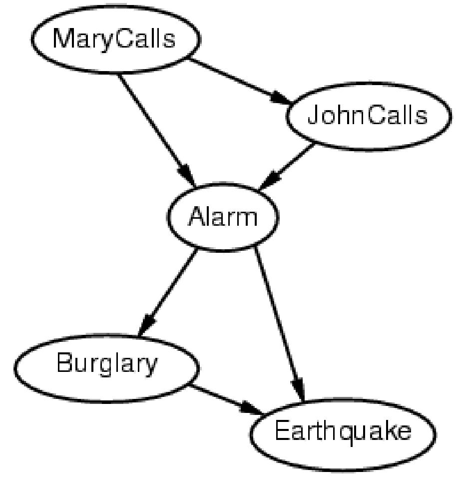
P(B | A, J, M) = P(B | A)? **Yes**

P(B | A, J, M) = P(B)? **No**

P(E | B, A ,J, M) = P(E | A)? **No**

P(E | B, A, J, M) = P(E | A, B)? **Yes**

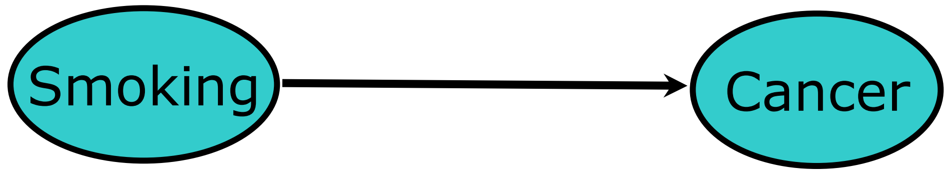
Example contd.



- Deciding conditional independence is hard in noncausal directions
-
- (Causal models and conditional independence seem hardwired for humans!)
-
- Network is less compact: $1 + 2 + 4 + 2 + 4 = 13$ numbers needed
-

Bayesian Networks

$$S \in \{no, light, heavy\}$$



$$C \in \{none, benign, malignant\}$$

P(S=no)	0.80
P(S=light)	0.15
P(S=heavy)	0.05

Smoking=	no	light	heavy
P(C=none)	0.96	0.88	0.60
P(C=benign)	0.03	0.08	0.25
P(C=malig)	0.01	0.04	0.15

Product Rule

- $P(C,S) = P(C|S)\,P(S)$

$S\Downarrow$ $C\Rightarrow$	*none*	*benign*	*malignant*
no	0.768	0.024	0.008
light	0.132	0.012	0.006
heavy	0.035	0.010	0.005

Marginalization

$S\Downarrow \quad C\Rightarrow$	*none*	*benign*	*malig*	total
no	0.768	0.024	0.008	.80
light	0.132	0.012	0.006	.15
heavy	0.035	0.010	0.005	.05
total	0.935	0.046	0.019	

P(Smoke)

P(Cancer)

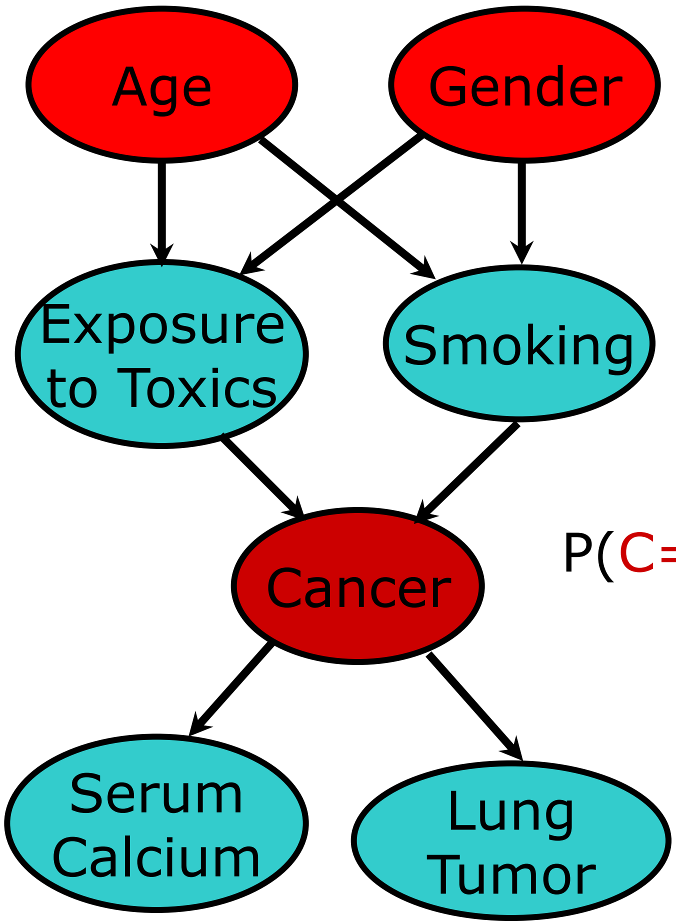
Bayes Rule Revisited

$$P(S \mid C) = \frac{P(C \mid S)P(S)}{P(C)} = \frac{P(C,S)}{P(C)}$$

$S\Downarrow \quad C\Rightarrow$	*none*	*benign*	*malig*
no	0.768/.935	0.024/.046	0.008/.019
light	0.132/.935	0.012/.046	0.006/.019
heavy	0.030/.935	0.015/.046	0.005/.019

Cancer=	none	benign	malignant
P(S=no)	0.821	0.522	0.421
P(S=light)	0.141	0.261	0.316
P(S=heavy)	0.037	0.217	0.263

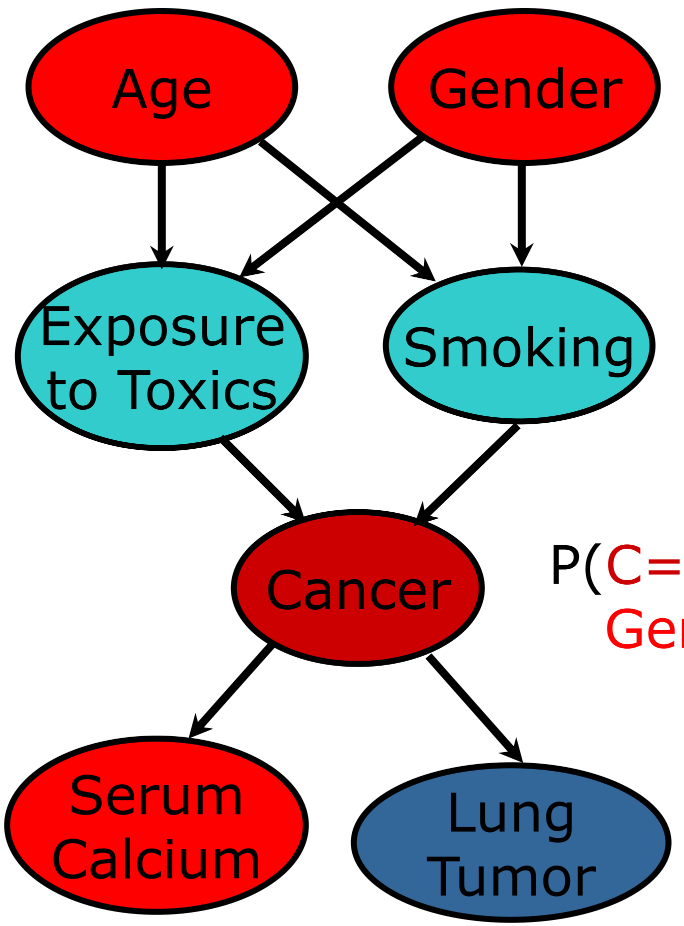
Predictive Inference



How likely are elderly males to get malignant cancer?

P(C=malignant | Age>60, Gender= male)

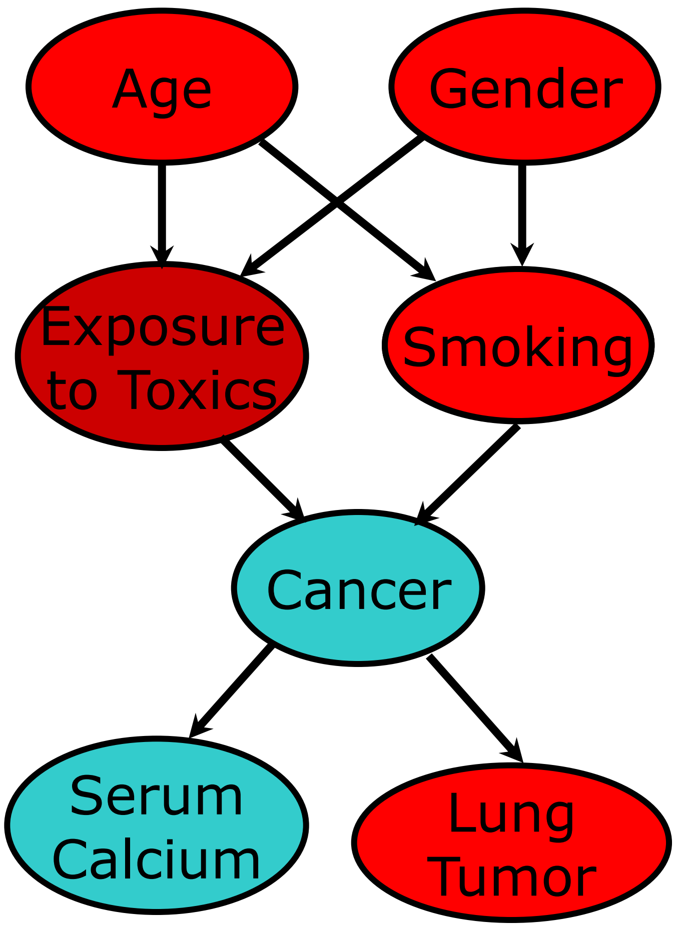
Combined



How likely is an elderly male patient with high Serum Calcium to have malignant cancer?

P(C=malignant | Age>60, Gender= male, Serum Calcium = high)

Explaining away



- If we see a lung tumor, the probability of heavy smoking and of exposure to toxics both go up.

- If we then observe heavy smoking, the probability of exposure to toxics goes back down.

Inference in Belief Networks

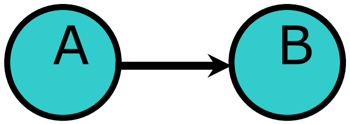
- Find $P(Q=q|\boldsymbol{E}= e)$
  - Q the query variable
  - $\boldsymbol{E}$ set of evidence variables

$$P(q \mid \mathbf{e}) = \frac{P(q, \mathbf{e})}{P(\mathbf{e})}$$

$X_1,\ldots, X_n$ are network variables except Q, **E**

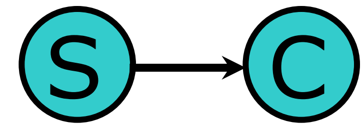
$$P(q, \mathbf{e}) = \sum_{x_1,\ldots,x_n} P(q, e, x_1,\ldots, x_n)$$

Basic Inference



$P(b) = ?$

Product Rule



- $P(C,S) = P(C|S)\,P(S)$

$S\Downarrow \quad C\Rightarrow$	*none*	*benign*	*malignant*
no	0.768	0.024	0.008
light	0.132	0.012	0.006
heavy	0.035	0.010	0.005

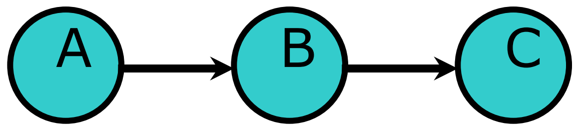
Marginalization

$S\Downarrow \quad C\Rightarrow$	*none*	*benign*	*malig*	total
no	0.768	0.024	0.008	.80
light	0.132	0.012	0.006	.15
heavy	0.035	0.010	0.005	.05
total	0.935	0.046	0.019	

P(Smoke)

P(Cancer)

Basic Inference



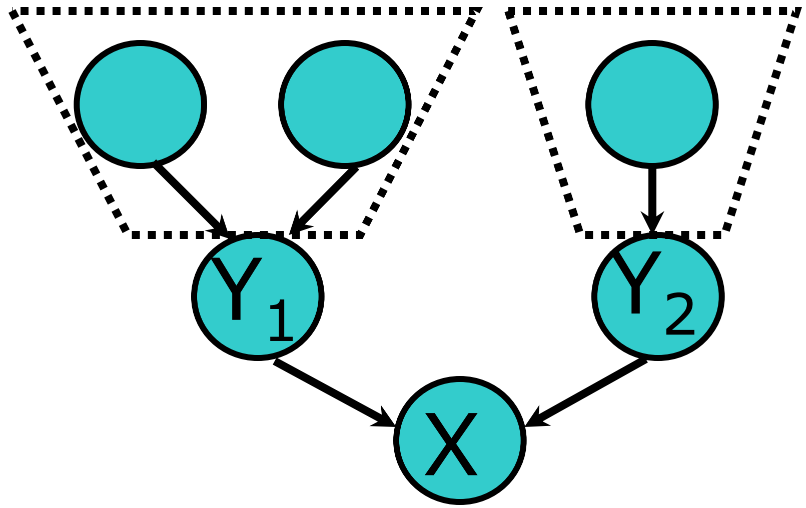
$$P(b) = \sum_a P(a, b) = \sum_a P(b \mid a)\, P(a)$$

$$P(c) = \sum_b P(c \mid b)\, P(b)$$

$$P(c) = \sum_{b,a} P(a, b, c) \qquad = \sum_{b,a} P(c \mid b)\, P(b \mid a)\, P(a)$$

$$= \sum_b P(c \mid b) \underbrace{\sum_a P(b \mid a)\, P(a)}_{P(b)}$$

Inference in trees



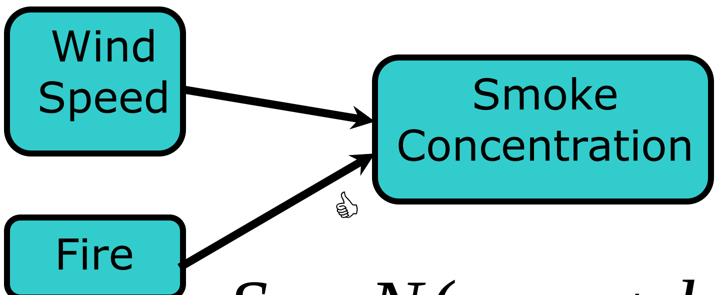
$$P(x) = \sum_{y_1, y_2} P(x \mid y_1, y_2)\, P(y_1, y_2)$$

because of independence of Y_1, Y_2:

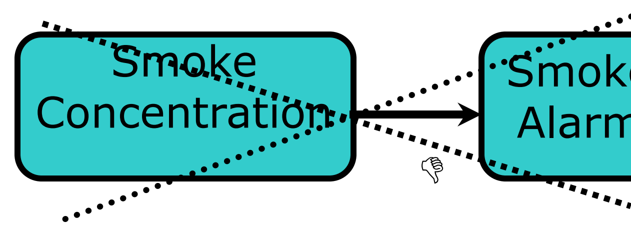
$$= \sum_{y_1, y_2} P(x \mid y_1, y_2)\, P(y_1)\, P(y_2)$$

Inference with continuous variables

- Gaussian networks: polynomial time inference regardless of network structure
- Conditional Gaussians:
  - discrete variables cannot depend on continuous



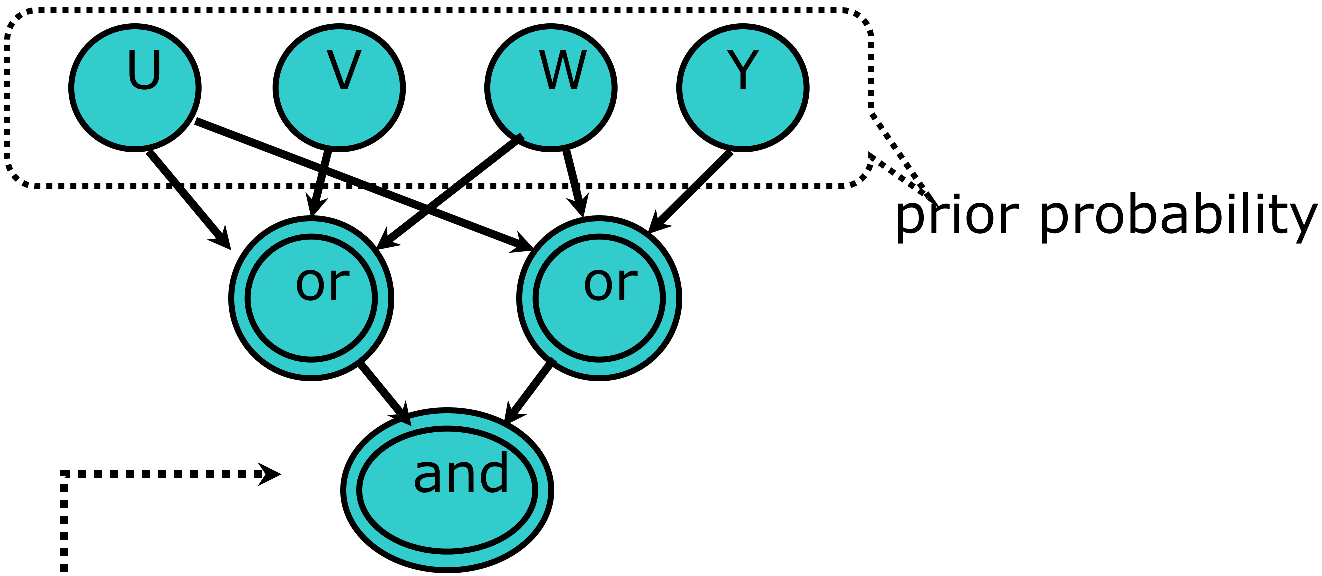
$$S \sim N(a_F w + b_F, \sigma_F^2)$$



- These techniques do not work for general hybrid networks.

- **Theorem:** Inference in a multi-connected Bayesian network is NP-hard.

Boolean 3CNF formula $\phi = (u \vee v \vee w) \wedge (u \vee w \vee y)$

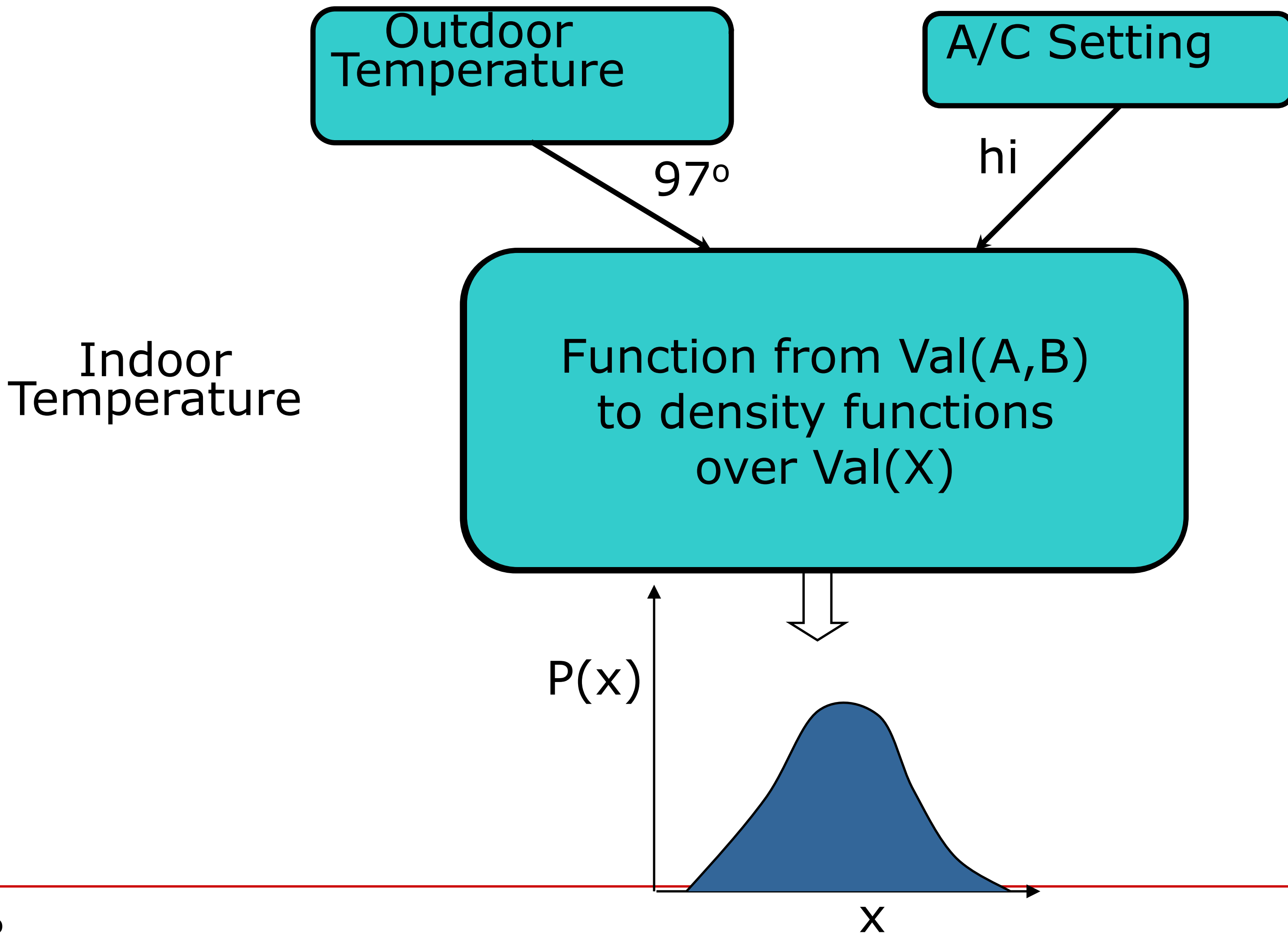


Probability ( ) = # satisfying assignments of ϕ

Summary

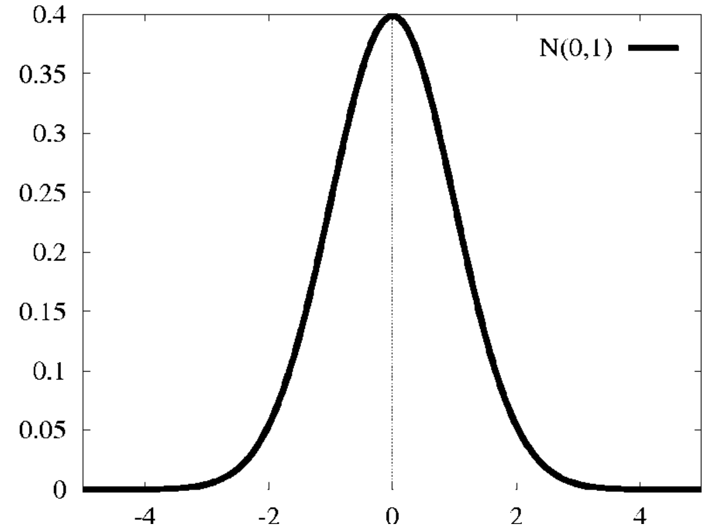
- Bayesian networks provide a natural representation for (causally induced) conditional independence
- Topology + CPTs = compact representation of joint distribution
- Generally easy for domain experts to construct

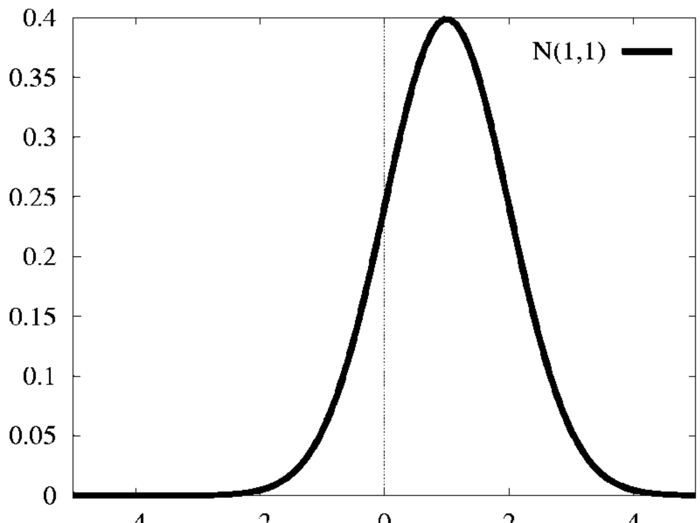
Continuous variables



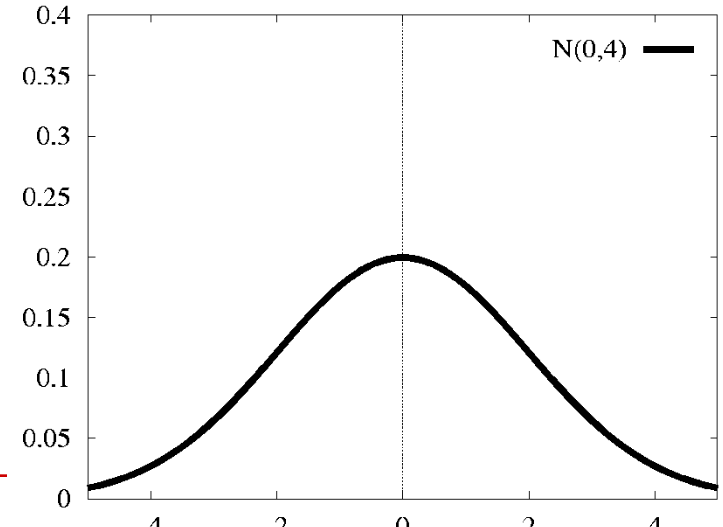
Gaussian (normal) distributions

$$\underbrace{P(x) = \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left(\frac{-(x-\mu)^2}{2\sigma}\right)}_{N(\mu,\ \sigma)}$$





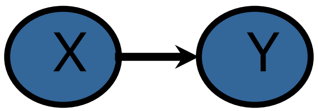
different mean



different variance

Gaussian networks

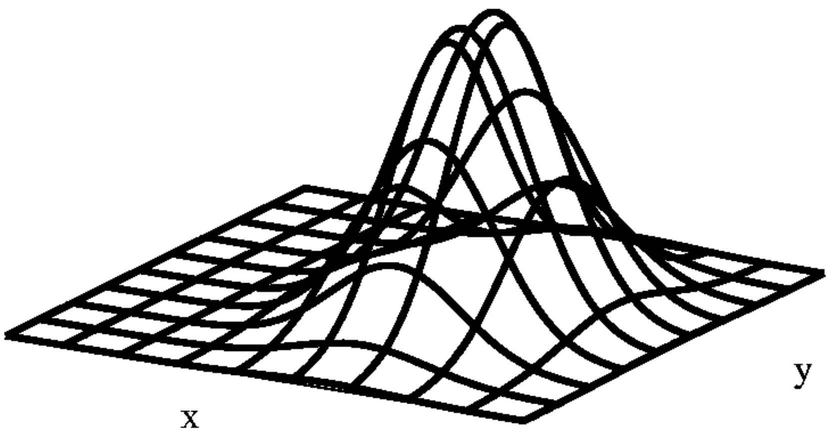
$$X \sim N(\mu, \sigma_X^2)$$

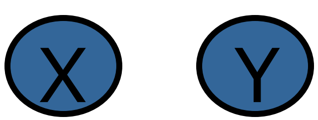


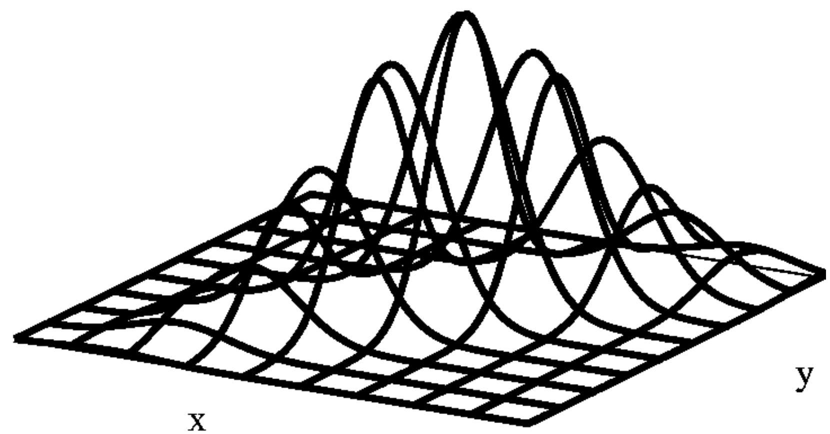
$$Y \sim N(ax + b, \sigma_Y^2)$$

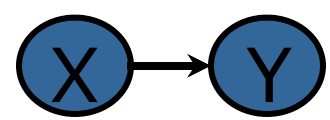
Each variable is a linear function of its parents, with Gaussian noise

Joint probability density functions:









D-Separation

- X is d-separated from Y if, **for all paths** between X and Y **there exists** an intermediate node Z for which:
  - The connection is serial or diverging and there is evidence for Z.
  - The connection is converging and Z (nor any of its descendants) have received any evidence.

- X is independent of Y given Z for some conditional probabilities if and only if X is d-separated from Y given Z.

Example of Independence Questions

- If there was evidence for B, which probabilities would change?
- If there was evidence for N, which probabilities would change?
- If there was evidence for M and N, which variables probabilities would change?
- Etc…