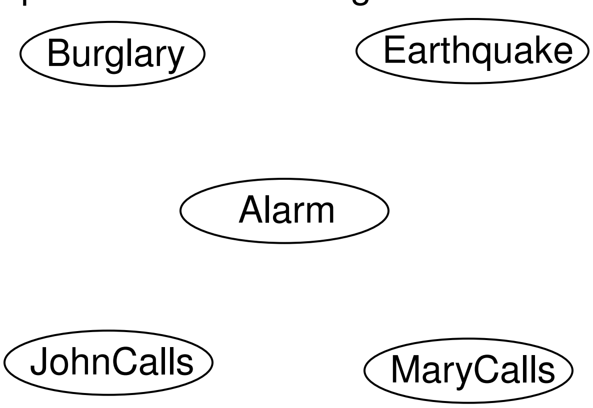
Bayes Nets
Representing and Reasoning about Uncertainty (Continued)

Combining the Two Examples

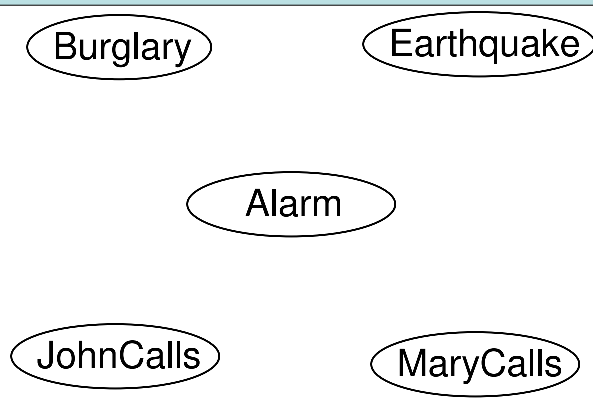
- I am at work, my neighbor John calls to say that my alarm went off, my neighbor Mary doesn't call. Sometimes the alarm is set off by a minor earthquake. Is there a burglar?



Earthquake Example

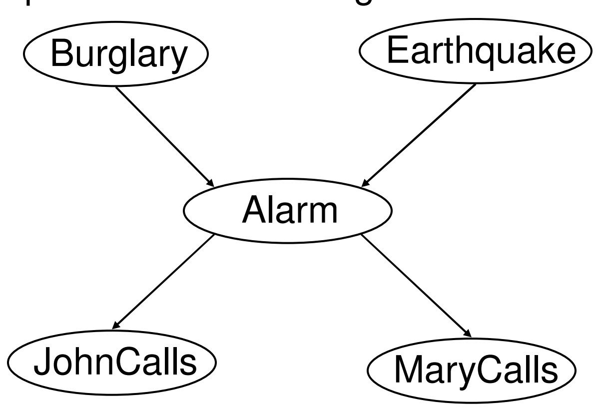
- I am at work, my neighbor John calls to say that my alarm went off, neighbor Mary doesn't call.

1: Define the variables that completely describe the problem.



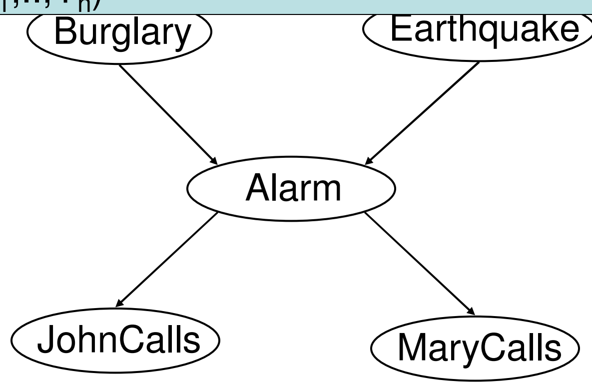
Earthquake Example

- I am at work, my neighbor John calls to say that my alarm went off, neighbor Mary doesn't call. Sometimes the alarm is set off by a minor earthquake. Is there a burglar?



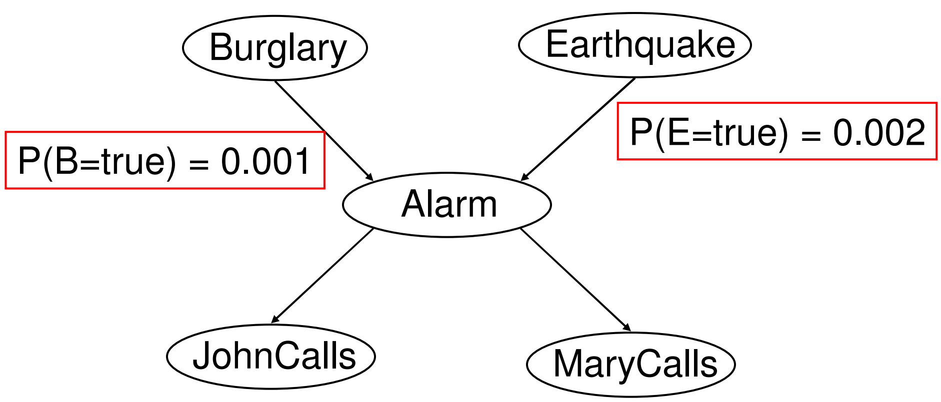
Earthquake Example

2: Define the links between variables.
• The resulting directed graph must be acyclic
• If node X has parents $Y_1,..,Y_n$, any variable that is not a descendent of X is conditionally independent of X given $(Y_1,..,Y_n)$



Earthquake Example

- I am at work, my neighbor John calls to say that my alarm went off, neighbor Mary doesn't call. Sometimes the alarm is set off by a minor earthquake. Is there a burglar?

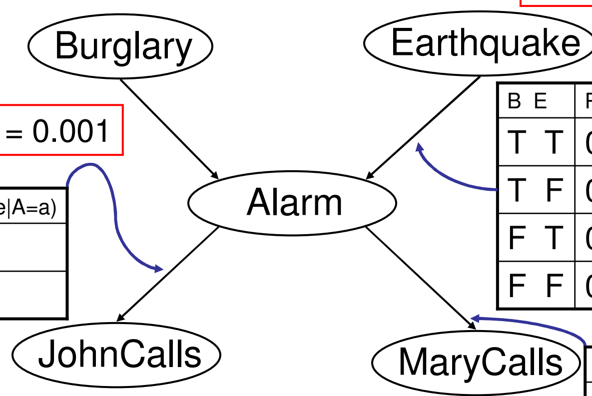


Earthquake Example

- I am at work, my neighbor John calls to say that my alarm went off, neighbor Mary doesn't call. Sometimes the alarm is set off by a minor earthquake. Is there a burglar?

P(E=true) = 0.002

P(B=True) = 0.001



B	E	P(A = True|B=b,E=e)
T	T	0.95
T	F	0.94
F	T	0.29
F	F	0.001

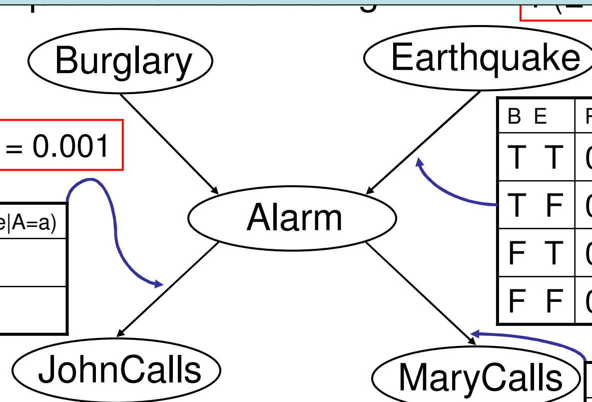
A	P(J = True|A=a)
T	0.90
F	0.05

A	P(M = True|A=a)
T	0.70
F	0.01

Earthquake Example

3: Add a probability table for each node. The table for node X contains P(X|Parent Values) for each possible combination of parent values

P(B=True) = 0.001



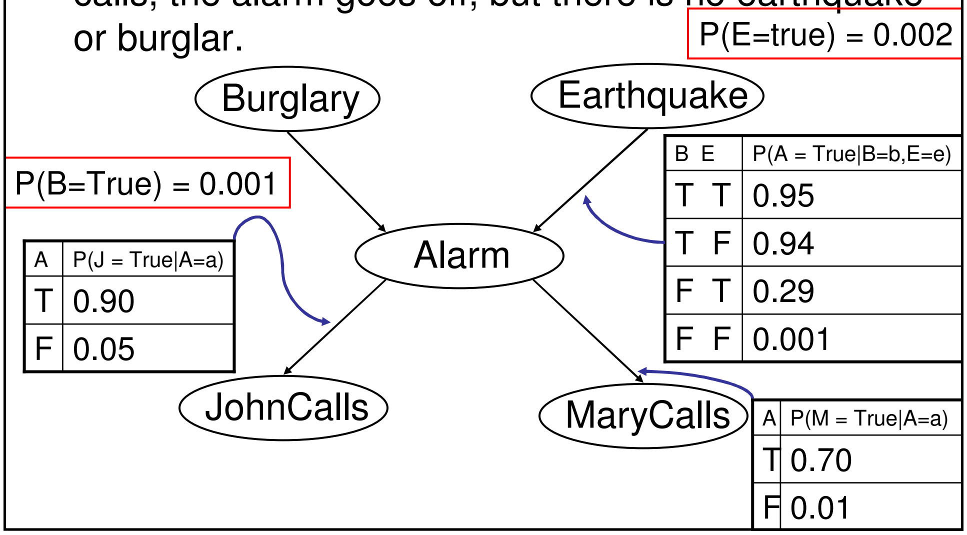
B	E	P(A = True|B=b,E=e)
T	T	0.95
T	F	0.94
F	T	0.29
F	F	0.001

A	P(J = True|A=a)
T	0.90
F	0.05

A	P(M = True|A=a)
T	0.70
F	0.01

Computing a Joint Entry

- *Any* entry in the joint probability table can be computed: Probability that both John and Mary calls, the alarm goes off, but there is no earthquake or burglar.



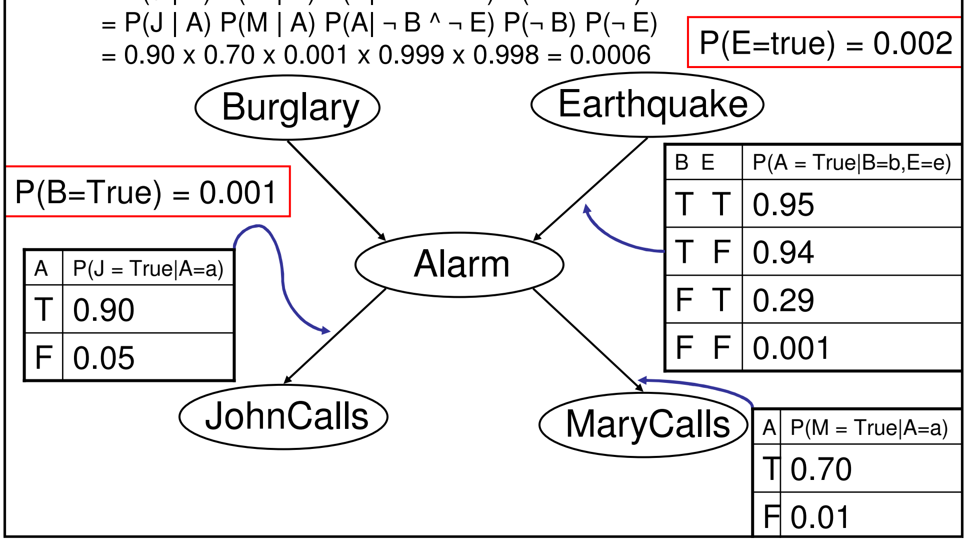
B	E	P(A = True|B=b,E=e)
T	T	0.95
T	F	0.94
F	T	0.29
F	F	0.001

A	P(J = True|A=a)
T	0.90
F	0.05

A	P(M = True|A=a)
T	0.70
F	0.01

Computing a Joint Entry

$$P(J \wedge M \wedge A \wedge \neg B \wedge \neg E) = P(J \mid M \wedge A \wedge \neg B \wedge \neg E)\, P(M \wedge A \wedge \neg B \wedge \neg E)$$
$$= P(J \mid A)\, P(M \wedge A \wedge \neg B \wedge \neg E)$$
$$= P(J \mid A)\, P(M \mid A \wedge \neg B \wedge \neg E)$$
$$= P(J \mid A)\, P(M \mid A)\, P(A \mid \neg B \wedge \neg E)\, P(\neg B \wedge \neg E)$$
$$= P(J \mid A)\, P(M \mid A)\, P(A \mid \neg B \wedge \neg E)\, P(\neg B)\, P(\neg E)$$
$$= 0.90 \times 0.70 \times 0.001 \times 0.999 \times 0.998 = 0.0006$$



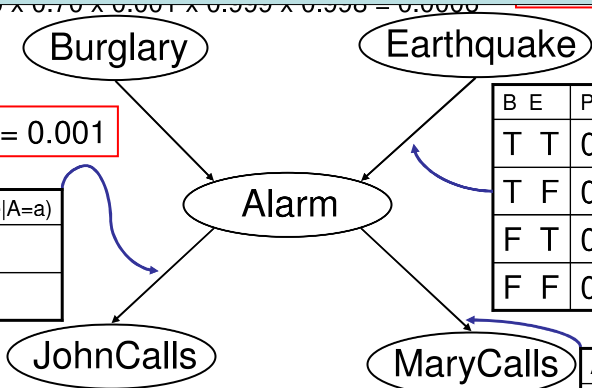
B	E	P(A = True|B=b,E=e)
T	T	0.95
T	F	0.94
F	T	0.29
F	F	0.001

A	P(J = True|A=a)
T	0.90
F	0.05

A	P(M = True|A=a)
T	0.70
F	0.01

Computing a Joint Entry

We would need 2^5 entries to store the entire joint distribution table.
But we need to store only 10 values by representing the dependencies between variables



P(B=True) = 0.001

B	E	P(A = True|B=b,E=e)
T	T	0.95
T	F	0.94
F	T	0.29
F	F	0.001

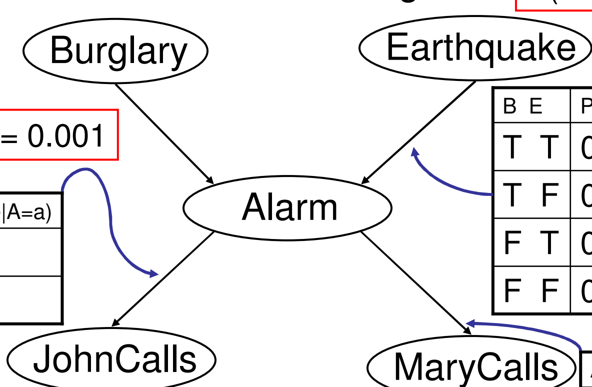
A	P(J = True|A=a)
T	0.90
F	0.05

A	P(M = True|A=a)
T	0.70
F	0.01

Inference

- *Any* inference operation of the form P(values of some variables | values of the other variables) can be computed: Probability that both John and Mary call given that there was a burglar.

P(E=true) = 0.002



P(B=True) = 0.001

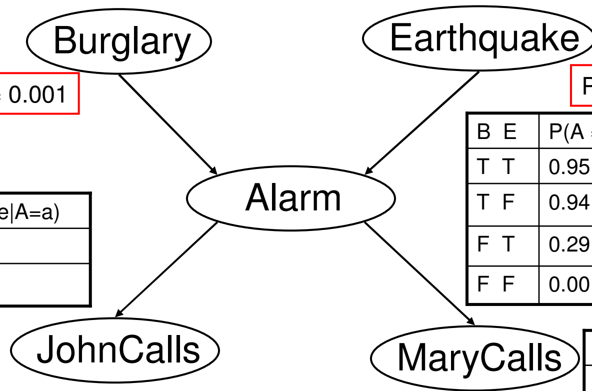
B	E	P(A = True|B=b,E=e)
T	T	0.95
T	F	0.94
F	T	0.29
F	F	0.001

A	P(J = True|A=a)
T	0.90
F	0.05

A	P(M = True|A=a)
T	0.70
F	0.01

Inference

$$P(J,M\mid B)=\frac{P(J,M,B)}{P(B)}=\frac{\sum_{\text{All entries } X \text{ that contain } J\wedge M\wedge B} P(X)}{\sum_{\text{All entries } Y \text{ that contain } B} P(Y)}$$



P(B=True) = 0.001

P(E=True) = 0.002

B	E	P(A = True|B=b,E=e)
T	T	0.95
T	F	0.94
F	T	0.29
F	F	0.001

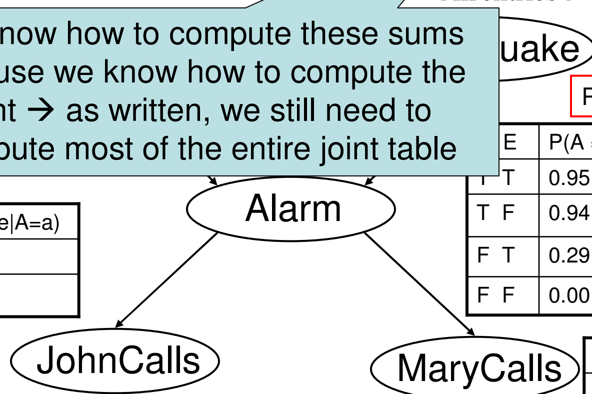
A	P(J = True|A=a)
T	0.90
F	0.05

A	P(M = True|A=a)
T	0.70
F	0.01

Inference

$$P(J,M\mid B)=\frac{P(J,M,B)}{P(B)}=\frac{\sum_{\text{All entries } X \text{ that contain } J\wedge M\wedge B} P(X)}{\sum_{\text{All entries } Y \text{ that contain } B} P(Y)}$$

We know how to compute these sums because we know how to compute the joint → as written, we still need to compute most of the entire joint table



P(

P(E=true) = 0.002

B	E	P(A = True|B=b,E=e)
T	T	0.95
T	F	0.94
F	T	0.29
F	F	0.001

A	P(J = True|A=a)
T	0.90
F	0.05

A	P(M = True|A=a)
T	0.70
F	0.01

Bayes Net: Formal Definition

- Bayes Net = directed acyclic graph represented by:
  - Set of vertices V
  - Set of directed edges E joining vertices. No cycles are allowed.
- With each vertex is associated:
  - The name of a random variable
  - A probability distribution table indicating how the probability of the variable’s values depends on all the possible combinations of values of its parents

Bayes Net: Formal Definition

- Bayes Net = directed acyclic graph represented by:
  - Set of vertices V
  - Set of directed edges E joining vertices. No cycles are allowed.
- With each vertex is associated:
  - The name of a random variable
  - A probability distribution table indicating how the probability of the variable’s values depends on all the possible combinations of values of its parents

Bayes Nets are also called Belief Networks
The tables associated with the vertices are called Conditional Probability Tables (CPT)
All the definitions can be extended to using continuous random variables instead of discrete variables

Bayes Net Construction

- Choose a set of variables and an ordering $\{X_1,..,X_m\}$
- For each variable X_i for i = 1 to m:
  1. Add the variable X_i to the network
  2. Set Parents(X_i) to be the minimal subset of $\{X_1,..,X_{i-1}\}$ such that X_i is conditionally independent of all the other members of $\{X_1,..,X_{i-1}\}$ given Parents(X_i)
  3. Define the probability table describing
     $P(X_i \mid Parents(X_i))$

Bayes Net Construction

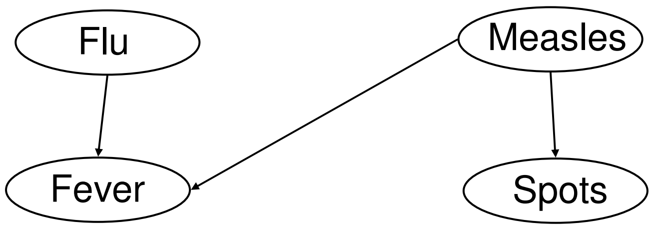
The structure of the network depends on the initial ordering of the variables

- Choose a set of variables and an ordering $\{X_1,..,X_m\}$
- For each variable X_i for i = 1 to m:
  1. Add the variable X_i to the network
  2. Set Parents(X_i) to be the minimal subset of $\{X_1,..,X_{i-1}\}$ such that X_i is conditionally independent of all the other members of $\{X_1,..,X_{i-1}\}$ given Parents(X_i)
  3. Define the probability table describing
     $P(X_i \mid Parents(X_i))$

If X_i has k parents, we need to store 2^k entries to represent the CPT → Storage is exponential in the number of parents, not in the total number of variables m. In many problems k << m.

Example: Symptoms & Diagnosis

- The diagnosis problem: What is the most likely disease given observed symptoms
- Variables V = {Flu, Measles, Fever, Spots}



Try creating the network by using a different ordering of the variables…..

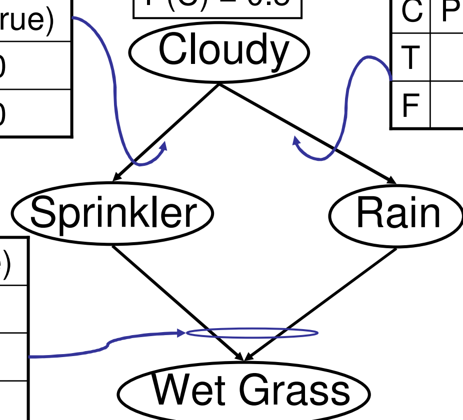
Another Examples

- The lawn may be wet because the sprinkler was on or because it was raining (or both).

P(C) = 0.5

C	P(S = True)
T	0.10
F	0.50

C	P(R = True)
T	0.80
F	0.20



S	R	P(W = True)
T	T	0.99
T	F	0.90
F	T	0.90
F	F	0.01

Computing the Joint: The General Case

$$
\begin{aligned}
&P(X_1 = x_1, X_2 = x_2, \ldots X_m = x_m) = \\
&P(X_m = x_m \mid X_1 = x_1, X_2 = x_2, \ldots X_{m-1} = x_{m-1}) \times \\
&P(X_1 = x_1, X_2 = x_2, \ldots X_{m-1} = x_{m-1}) = \\
&P(X_m = x_m \mid X_1 = x_1, X_2 = x_2, \ldots X_{m-1} = x_{m-1}) \times \\
&P(X_{m-1} = x_{m-1} \mid X_1 = x_1, X_2 = x_2, \ldots X_{m-2} = x_{m-2}) \times \\
&P(X_1 = x_1 . X_2 = x_2, \ldots X_{m-2} = x_{m-2}) = \\
&\vdots \\
&\prod_{i=1}^{m} P(X_i = x_i \mid X_1 = x_1, X_2 = x_2, \ldots X_{i-1} = x_{i-1}) = \\
&\boxed{\prod_{i=1}^{m} P(X_i = x_i \mid \text{assignments to Parents}(X_i))}
\end{aligned}
$$

- Any entry in the joint distribution table can be computed
- Consequently, any conditional probability can be computed

Computing the Joint: The General Case

$$
\begin{aligned}
&P(X_1 = x_1, X_2 = x_2, \ldots X_m = x_m) = \\
&P \quad {}_{m-1}) \times \\
&P \\
&P \quad {}_{m-1}) \times \\
&P(X_{m-1} = x_{m-1} \mid X_1 = x_1, X_2 = \quad {}_{m-2} = x_{m-2}) \times \\
&P(X_1 = x_1 . X_2 = x_2, \ldots X_{m-2} = x_{m-2} \\
&\vdots \\
&\prod_{i=1}^{m} P(X_i = x_i \mid X_1 = x_1, X_2 = x_2, \ldots X_{i-1} = x_{i-1}) = \\
&\boxed{\prod_{i=1}^{m} P(X_i = x_i \mid \text{assignments to Parents}(X_i))}
\end{aligned}
$$

We can do this because, by construction, X_i is independent of all the other variables given Parents(X_i)

- Any entry in the joint distribution table can be computed
- Consequently, any conditional probability can be computed

Inference: The General Case

- Inference = Computing a conditional probability:

P(Value for some variable(s) | Values for other variables)

Inference: The General Case

- Inference = Computing a conditional probability:

P(Value for some variable(s) | Values for other variables)

"Query" variables
Example: Disease

"Evidence" variables
Example: Symptoms

$$P(E_1 \mid E_2)$$

Also called "belief updating"

Inference: The General Case

- Inference = Computing a conditional probability:

P(Value for some variable(s) | Values for other variables)

$$\boldsymbol{P}(\boldsymbol{E}_1 \mid \boldsymbol{E}_2) = \frac{\boldsymbol{P}(\boldsymbol{E}_1, \boldsymbol{E}_2)}{\boldsymbol{P}(\boldsymbol{E}_2)} = \frac{\displaystyle\sum_{\text{All joint entries } \boldsymbol{X} \text{ that contain } \boldsymbol{E}_1 \wedge \boldsymbol{E}_2} \boldsymbol{P}(\boldsymbol{X})}{\displaystyle\sum_{\text{All joint entries } \boldsymbol{Y} \text{ that contain } \boldsymbol{E}_2} \boldsymbol{P}(\boldsymbol{Y})}$$

We can compute any conditional probability so we can perform solve any inference problem in principle

So Far…

- Methodology for building Bayes nets.
- Requires exponential storage in the maximum number of parents of any node, not in the total number of nodes.
- We can compute the value of any assignment to the variables (entry in the joint distribution) in time linear in the number of variables.
- We can compute the answer to any question (any conditional probability)

Inference: The General Case

Problem: if E_2 involves k binary variables and we have a total of m variables, what is the complexity of this computation?

a conditional

values for other variables)

$$P(E_1 \mid E_2) = \frac{P(E_1, E_2)}{P(E_2)} = \frac{\sum_{\text{All joint entries } X \text{ that contain } E_1 \wedge E_2} P(X)}{\sum_{\text{All joint entries } Y \text{ that contain } E_2} P(Y)}$$

We can compute any conditional probability so we can perform solve any inference problem in principle

Inference: The Bad News

- Computing the conditional probabilities by enumerating all relevant entries in the joint is expensive:

Exponential in the number of variables!

- Even worse:

Solving for general queries in Bayes nets is NP-hard!

Possible Solutions

- Approximate methods
  - Approximate the joint distributions by drawing samples
- Exact methods
  - Factorization and variable elimination
  - Exploit special network structure (e.g., trees)
  - Transform the network structure

Approximate Method: Sampling

- Sampling = Very powerful technique in many probabilistic problems
- General idea:
  - It is often difficult to compute and represent exactly the probability distribution of a set of variables
  - But, it is often easy to generate examples from the distribution

The number of rows too large for the table to be computed explicitly

x_1 $x_{2...}x_m$	$P(X_1=x_1,X_2=x_2,\ldots,X_m = x_m)$
T T…T	0.95
T F…T	0.94
F T…T	0.29
F F…T	0.001
…….	……………….

Approximate Method: Sampling

- Sampling = Very powerful technique in many probabilistic problems (stochastic simulation)
- General idea:
  - It is often difficult to compute and represent exactly the probability distribution of a set of variables
  - But, it is often easy to generate examples from the distribution

P ➡

x_1 x_2.......................x_m
T T F T F F…T F
T F T T T F…T T
F T T F F F…T F
…….
F F F T F T…F T

For a large number of samples,
$P(X_1=x_1, X_2=x_2, \ldots, X_m = x_m)$
is approximately equal to:

$$\frac{\text{\# of samples with } X_1=x_1 \text{ and } X_2=x_2 \ldots \text{and } X_m = x_m}{\text{Total \# of samples}}$$

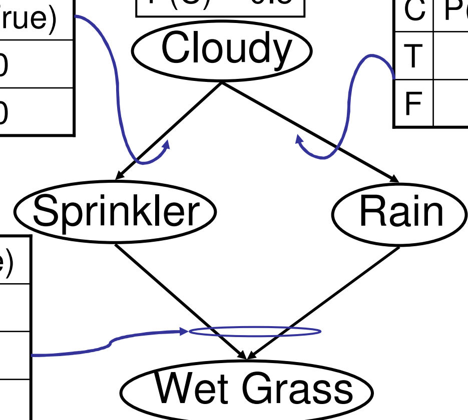
Sampling Example

- Generate a set of variable assignments with the same distribution as the joint distribution represented by the network

P(C) = 0.5

C	P(S = True)
T	0.10
F	0.50

C	P(R = True)
T	0.80
F	0.20



S	R	P(W = True)
T	T	0.99
T	F	0.90
F	T	0.90
F	F	0.01

Sampling

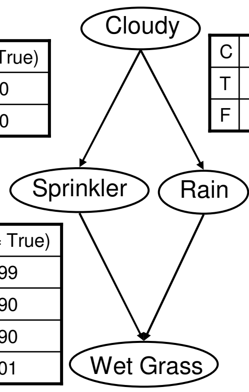
C	S	R	W
T			

1. Randomly choose C.
   C = True with probability 0.5
   → C = True

P(C) = 0.5

C	P(S = True)
T	0.10
F	0.50

C	P(R = True)
T	0.80
F	0.20



S	R	P(W = True)
T	T	0.99
T	F	0.90
F	T	0.90
F	F	0.01

Sampling

C	S	R	W
T	F		

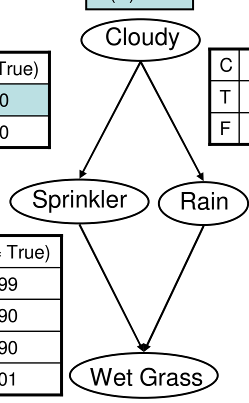
1. Randomly choose C.
   C = True with probability 0.5
   → C = True

2. Randomly choose S.
   S = True with probability 0.10
   → S = False

P(C) = 0.5

C	P(S = True)
T	0.10
F	0.50

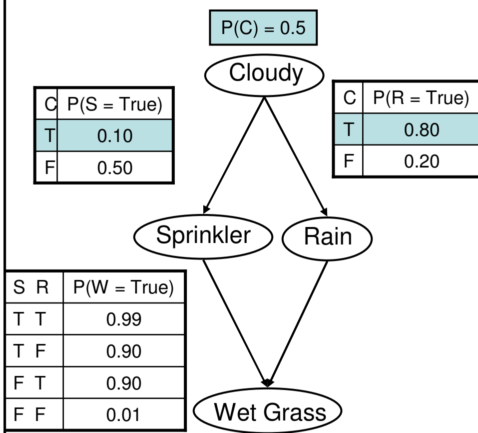
C	P(R = True)
T	0.80
F	0.20



S	R	P(W = True)
T	T	0.99
T	F	0.90
F	T	0.90
F	F	0.01

Sampling

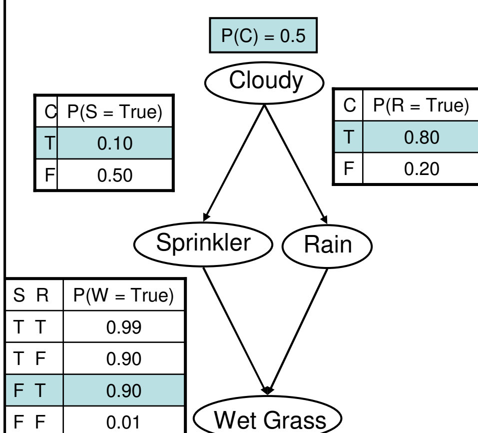
C	S	R	W
T	F	T	



1. Randomly choose C.
   C = True with probability 0.5
   → C = True
2. Randomly choose S.
   S = True with probability 0.10
   → S = False
3. Randomly choose R.
   R = True with probability 0.80
   → R = True

Sampling

C	S	R	W
T	F	T	T



1. Randomly choose C.
   C = True with probability 0.5
   → C = True
2. Randomly choose S.
   S = True with probability 0.10
   → S = False
3. Randomly choose R.
   R = True with probability 0.80
   → R = True
4. Randomly choose W.
   W = True with probability 0.90
   → W = True

Sampling for Inference: Example

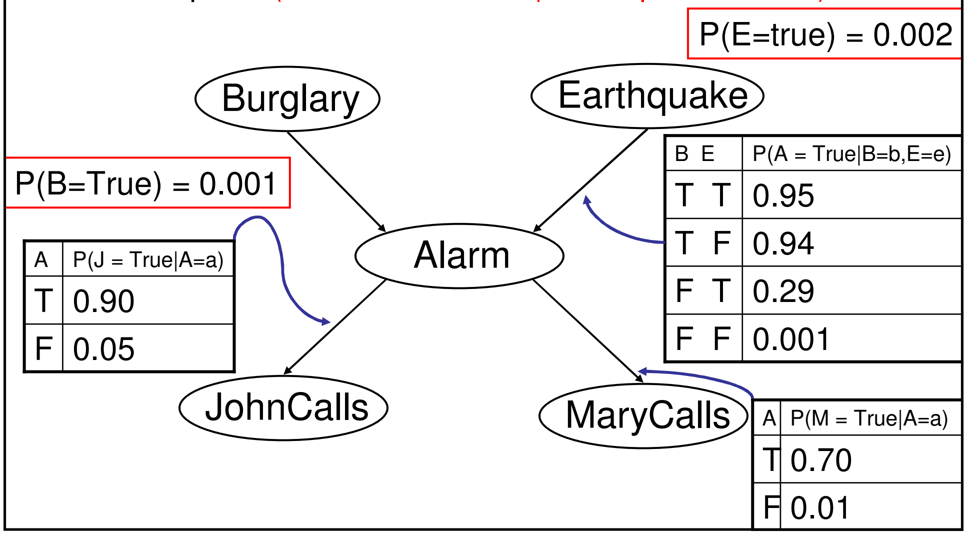
- Suppose that we want to compute P(W = True | C = True) (In words: How likely is it that the grass will be wet given that the sky is cloudy)
- Compute lots of samples of (C,S,R,W)
  - N_c = Number of samples for which C = True
  - N_s = Number of samples for which W = True and C = True
  - N = Total number of samples
- N_c/N approximates P(C = True)
- N_s/N approximates P(W = True and C = True)
- Therefore: N_s/N_c approximates: P(W = True and C = True)/ P(C = True) = P(W = True | C = True)

Sampling for Inference: General Case

- Suppose that we want to compute $P(E_1| E_2)$ (In words: How likely is it that the variable assignments in E_1 are satisfied given the assignments in E_2)
- Compute lots of samples
  - N_c = Number of samples for which the assignments in E_2 are satisfied
  - N_s = Number of samples for which the assignments in E_1 are satisfied
  - N = Total number of samples
- N_c/N approximates $P(E_2)$
- N_s/N approximates $P(E_1 \text{ and } E_2)$
- Therefore: N_s/N_c approximates: $P(E_1 \text{ and } E_2)/ P(E_2) = P(E_1|E_2)$

Problem with Sampling

- Probability is so low for some assignments of variables that that will likely never be seen in the samples (unless a very large number of samples is drawn).
- Example: P(JohnCalls = True | Earthquake = True)



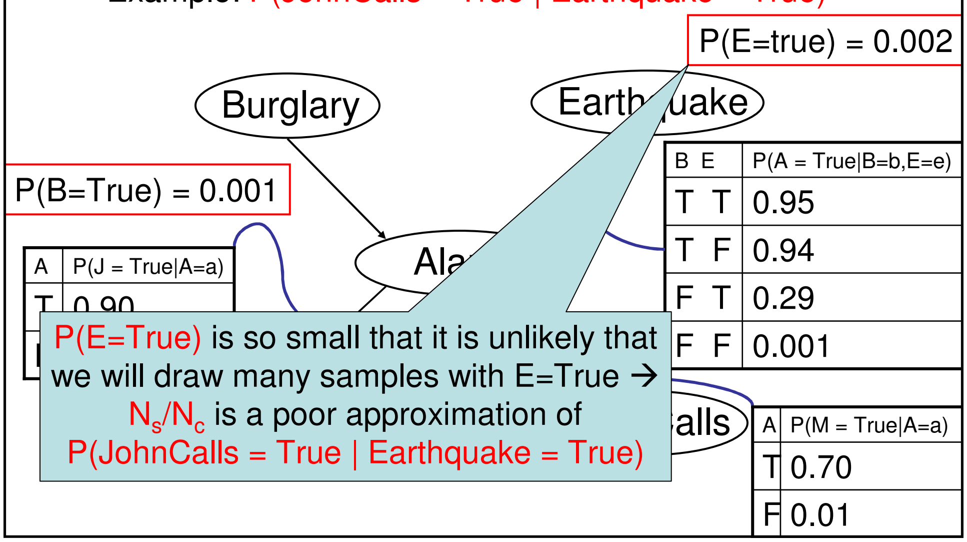
B	E	P(A = True|B=b,E=e)
T	T	0.95
T	F	0.94
F	T	0.29
F	F	0.001

A	P(J = True|A=a)
T	0.90
F	0.05

A	P(M = True|A=a)
T	0.70
F	0.01

Problem with Sampling

- Probability is so low for some assignments of variables that that they will likely never be seen in the samples (unless a very large number of samples is drawn).
- Example: P(JohnCalls = True | Earthquake = True)



B	E	P(A = True|B=b,E=e)
T	T	0.95
T	F	0.94
F	T	0.29
F	F	0.001

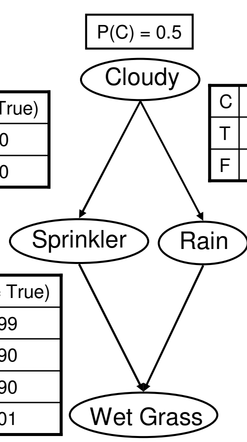
A	P(J = True|A=a)
T	0.90

A	P(M = True|A=a)
T	0.70
F	0.01

Solution: Likelihood Weighting

- Suppose that E_2 contains a variable assignment of the form $X_i = v$
- Current approach:
  - Generate samples until enough of them contain $X_i = v$
  - Such samples are generated with probability $p = P(X_i = v \mid \text{Parents}(X_i))$
- Likelihood Weighting:
  - Generate only samples with $X_i = v$
  - Weight each sample by $\omega = p$

Likelihood Weighting



C	P(S = True)
T	0.10
F	0.50

C	P(R = True)
T	0.80
F	0.20

S	R	P(W = True)
T	T	0.99
T	F	0.90
F	T	0.90
F	F	0.01

Example: Suppose that we want to compute an inference with

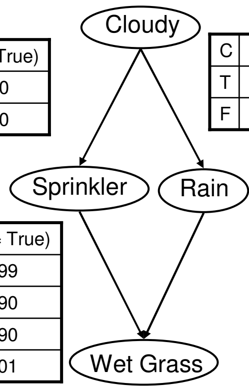
E_2 = (Sprinkler = True , Wet Grass = True)

Likelihood Weighting

$\omega = 1.0$

1. Randomly choose C.
   C = True with probability 0.5
   → C = True

P(C) = 0.5



C	P(S = True)
T	0.10
F	0.50

C	P(R = True)
T	0.80
F	0.20

S	R	P(W = True)
T	T	0.99
T	F	0.90
F	T	0.90
F	F	0.01

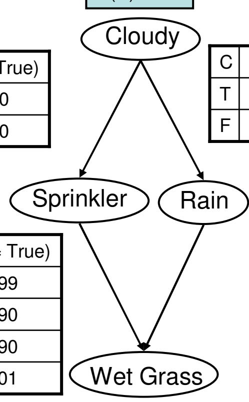
Likelihood Weighting

$\omega = 1.0$

1. Randomly choose C.
   C = True with probability 0.5
   → C = True

C is not one of the evidence variables, so we take a random sample as before

P(C) = 0.5



C	P(S = True)
T	0.10
F	0.50

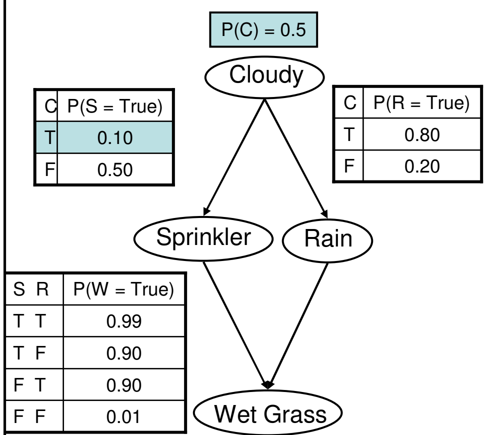
C	P(R = True)
T	0.80
F	0.20

S	R	P(W = True)
T	T	0.99
T	F	0.90
F	T	0.90
F	F	0.01

Likelihood Weighting

$\omega = 1.0 \times 0.10$

1. Randomly choose C.
   C = True with probability 0.5
   → C = True
2. Set S = True



C	P(S = True)
T	0.10
F	0.50

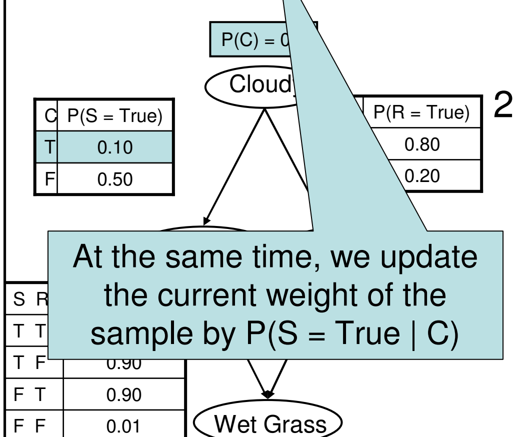
C	P(R = True)
T	0.80
F	0.20

S R	P(W = True)
T T	0.99
T F	0.90
F T	0.90
F F	0.01

Likelihood Weighting

$\omega = 1.0 \times 0.10$

1. Randomly choose C.
   C = True with probability 0.5
   → C = True
2. Set S = True



C	P(S = True)
T	0.10
F	0.50

	P(R = True)
	0.80
	0.20

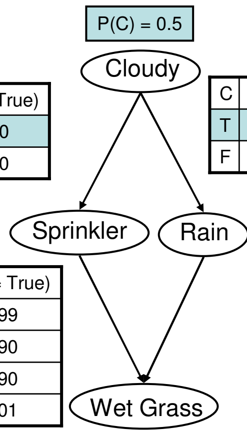
S R	
T T	
T F	0.90
F T	0.90
F F	0.01

At the same time, we update the current weight of the sample by P(S = True | C)

S is one of the evidence variables, so we fix its value *without* sampling

Likelihood Weighting

$\omega = 1.0 \times 0.10$



C	P(S = True)
T	0.10
F	0.50

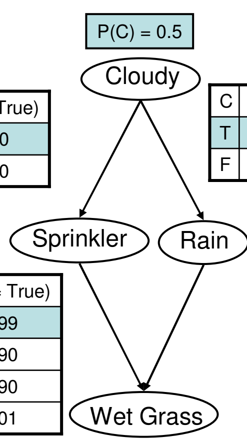
C	P(R = True)
T	0.80
F	0.20

S	R	P(W = True)
T	T	0.99
T	F	0.90
F	T	0.90
F	F	0.01

1. Randomly choose C.
   C = True with probability 0.5
   → C = True
2. Set S = True
3. Randomly choose R.
   R = True with probability 0.80
   → R = True

Likelihood Weighting

$\omega = 1.0 \times 0.10 \times 0.99$



C	P(S = True)
T	0.10
F	0.50

C	P(R = True)
T	0.80
F	0.20

S	R	P(W = True)
T	T	0.99
T	F	0.90
F	T	0.90
F	F	0.01

1. Randomly choose C.
   C = True with probability 0.5
   → C = True
2. Set S = True
3. Randomly choose R.
   R = True with probability 0.80
   → R = True
4. Set W = True

Likelihood Weighting

ω = 1.0 x 0.10 x 0.99

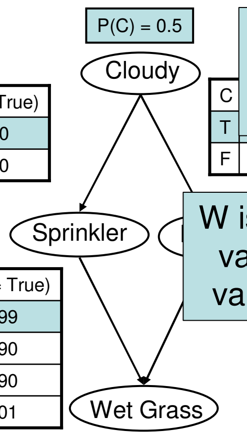
At the same time, we update the current weight of the sample by P(W = True | S,R)

W is one of the evidence variables, so we fix its value *without* sampling

C	P(S = True)
T	0.10
F	0.50

C	P(R = True)
T	
F	0.20

S R	P(W = True)
T T	0.99
T F	0.90
F T	0.90
F F	0.01



1. Randomly choose C.
   C = True with probability 0.5
   → C = T...

...ly choose R.
...e with probability
... = True

4. Set W = True

Likelihood Weighting

C	S	R	W
T	T	T	T

w = 0.099

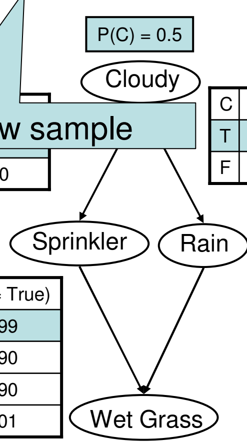
New sample

Weight associated with the sample. We increment the "number" of samples by ω:
$N_c \leftarrow N_c + \omega$

C	P(S = True)
T	
F	0.50

C	P(R = ...
T	0.80
F	0.20

S R	P(W = True)
T T	0.99
T F	0.90
F T	0.90
F F	0.01



1. Randomly choose C.
   C = True with probability 0.5
   → C = True

2. Set S = True

...ndomly choose R.
...bility

4. Set W = True

Likelihood Weighting

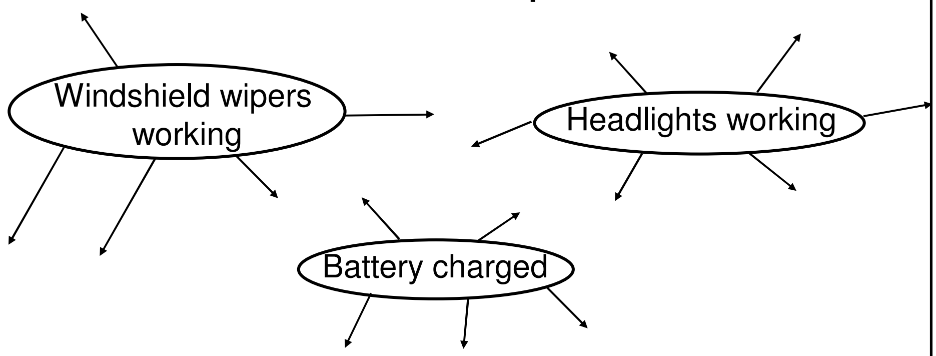
- $N_c = 0$; $N_s = 0$;
  1. Generate a random assignment of the variables, fixing the variables assigned in E_2
  2. Assign the sample a weight ω = probability that this sample would have been generated if we did not fix the value of the variables in E_2
  3. $N_c \leftarrow N_c + \omega$
  4. If the sample matches E_1 $N_s \leftarrow N_s + \omega$
  5. Repeat until we have "enough" samples
- N_s/N_c is an estimate of $P(E_1|E_2)$

Possible Solutions

- Approximate methods
  - Approximate the joint distributions by drawing samples
- Exact methods
  - Factorization and variable elimination
  - Exploit special network structure (e.g., trees)
  - Transform the network structure

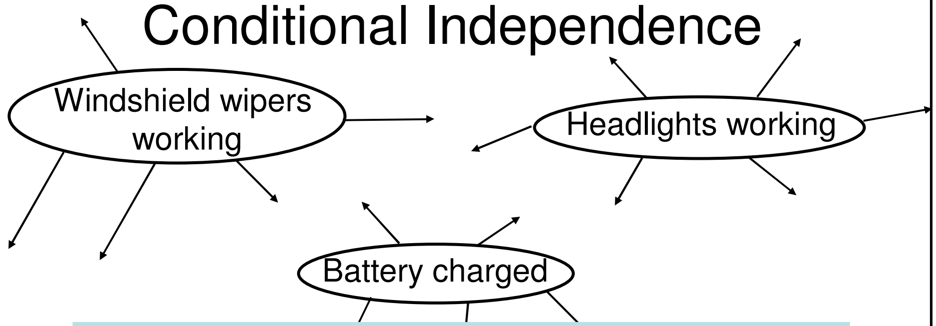
Before that, let's first look at other independence information that can be extracted

More Powerful Statements about Conditional Independence



Assume that these nodes are part of a much larger network and may be very far away from each other (slightly contrived example)

More Powerful Statements about Conditional Independence



For this type of engine, once we know B, the knowledge of W is irrelevant to H.
H and W are conditionally independent given B:

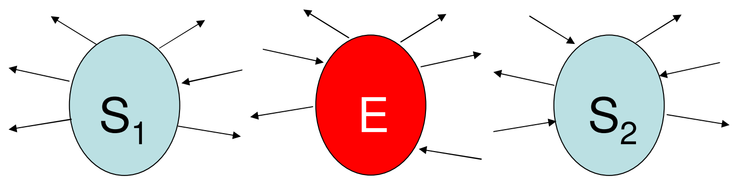
$$P(H|W,B) = P(H|B)$$

How can we find out this independence relation?

More General

- Given 2 sets of nodes S_1 and S_2
- Given a set of nodes E to which we have assigned values (the evidence set)
- Are S_1 and S_2 conditionally independent given E?

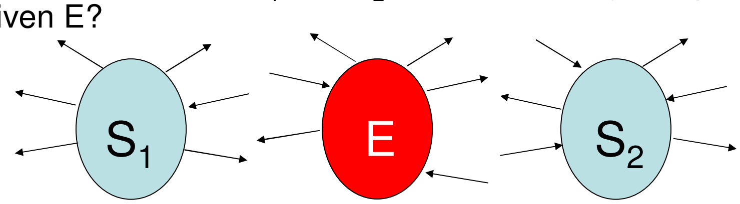
P (assignments to S_1 | E and assignments to S_2) = P (assignments to S_1 | E)



- Why is it important and useful?

More General

- How can we find if S_1 and S_2 are conditionally independent given E?



P (assignments to S_1 | E and assignments to S_2) = P (assignments to S_1 | E)

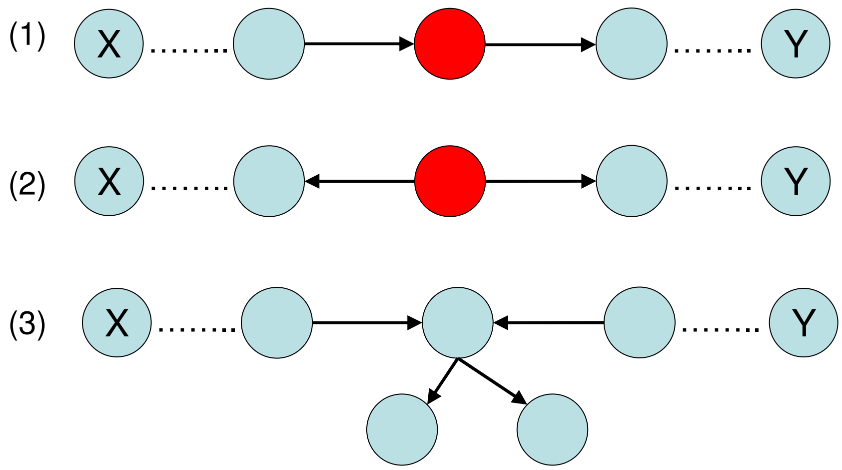
- Why is it important and useful?

We can simplify any computation that contains something like $P(S_1 \mid E, S_2)$ by $P(S_1 \mid E)$

Intuitively E stands in between or “blocks” S_1 from S_2

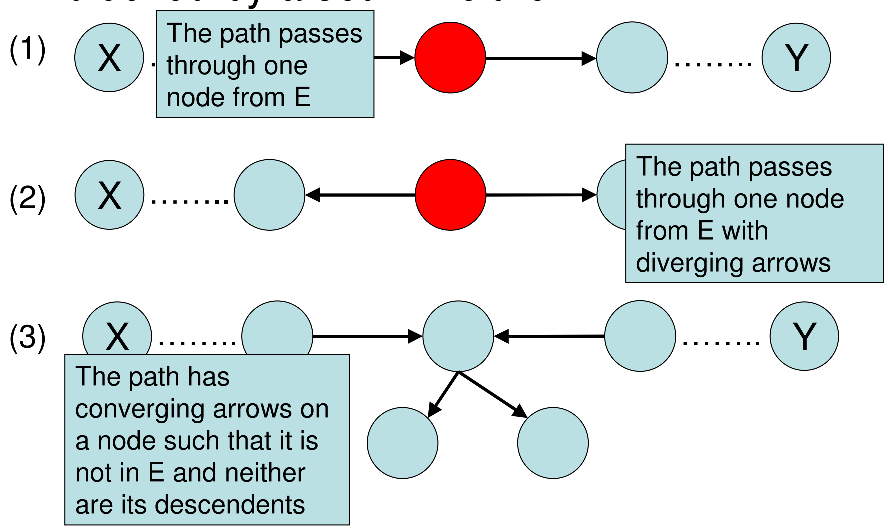
Blockage: Formal Definition

- A path from a node X to a node Y is *blocked* by a set E if either:



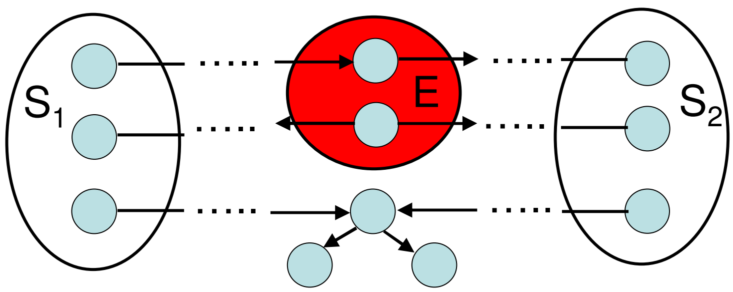
Blockage: Formal Definition

- A path from a node X to a node Y is *blocked* by a set E if either:



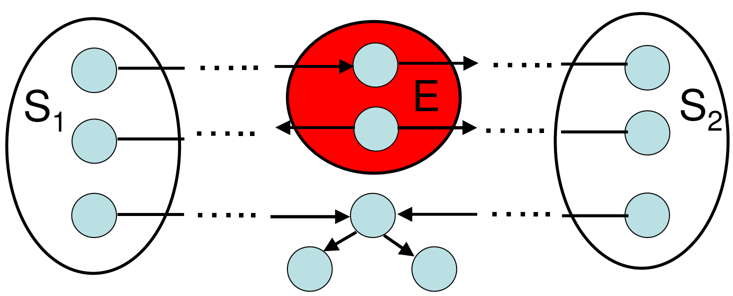
D-Separation Theorem

- If every (undirected) path from a node in a set S_1 to one in a set S_2 is blocked by E, then E *d-separates* S_1 and S_2.

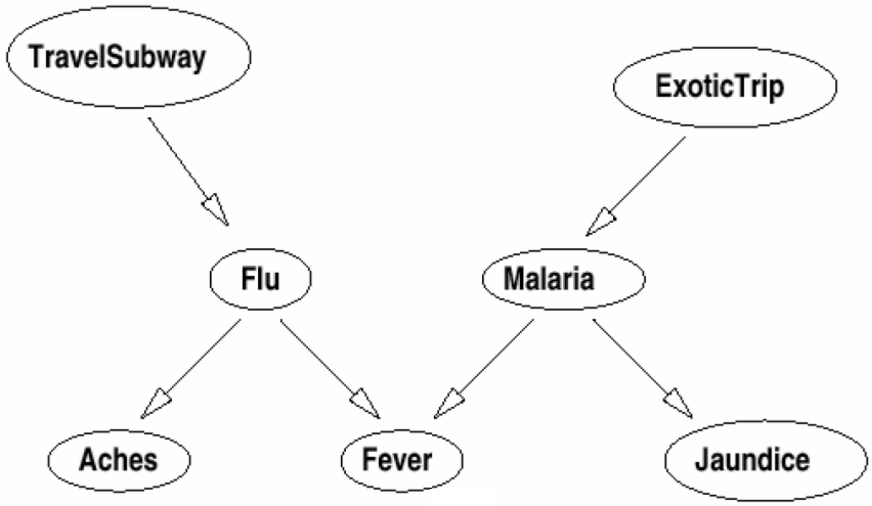


D-Separation Theorem

- If E *d-separates* S_1 and $S_{2,}$ then S_1 and S_2 are conditionally independent given E.
  - $P(S_2 \mid S_1 , E) = P(S_2 \mid E)$
  - $P(S_1 \mid S_2 , E) = P(S_1 \mid E)$



Aches and Fever?
Flu and Malaria?
Subway and ExoticTrip?



So Far…

- Methodology for building Bayes nets.
- Requires exponential storage in the maximum number of parents of any node.
- We can compute the value of any assignment to the variables (entry in the joint distribution) in time linear in the number of variables.
- We can compute the answer to any question (any conditional probability)
- But inference is NP-hard in general
- Sampling (stochastic sampling) for approximate inference
- D-separation criterion to be used for extracting additional independence relations (and simplifying inference)