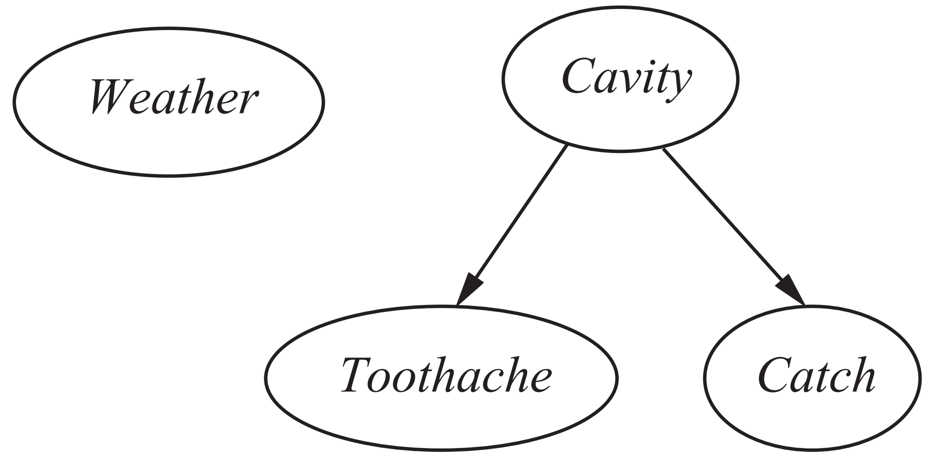
Bayes Net

CSEN266

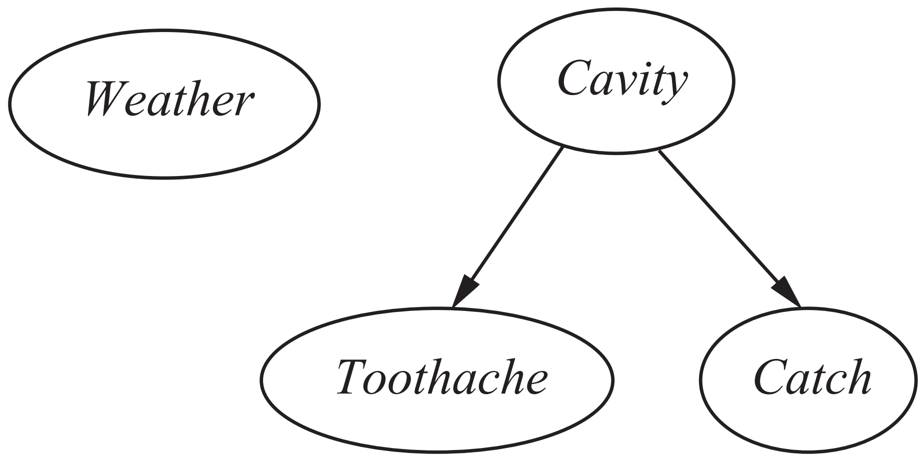
Artificial Intelligence

Bayesian Net



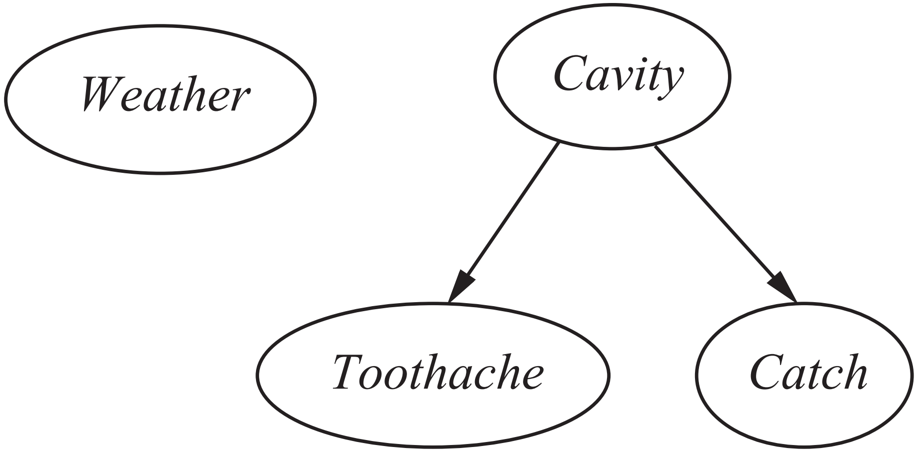
- Causal relation among several events
- Observe some event(s)
- Want to infer the probability of other events

Bayesian Net



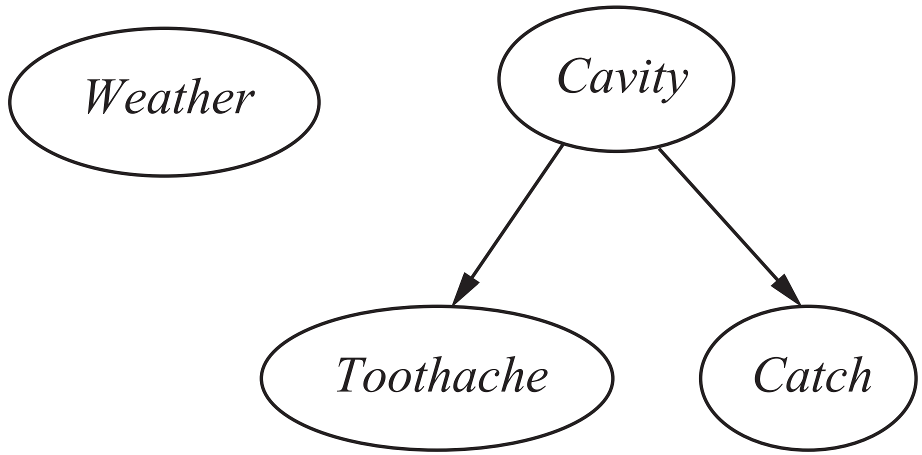
- Four variables:
  - $Cavity$: a direct cause of $Toothache$ and $Catch$
  - $Toothache$ and $Catch$: no direct causal relationship
  - $Toothache$ and $Catch$: conditionally independent, given $Cavity$

Bayesian Net



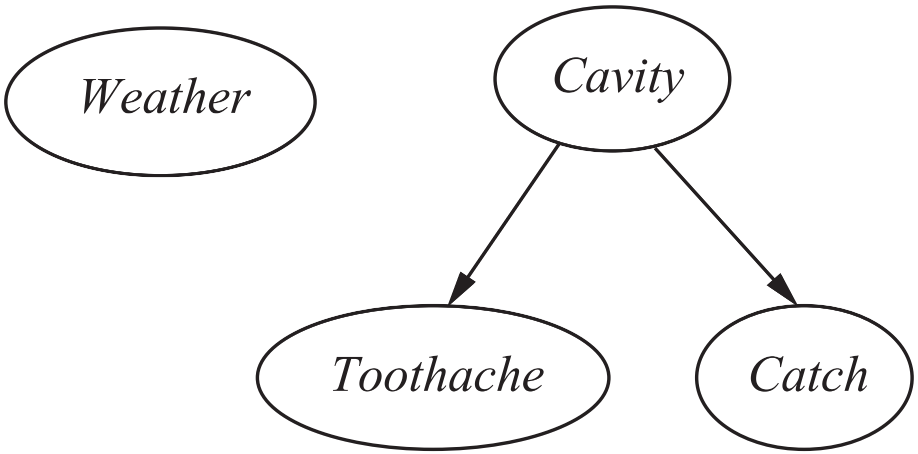
- Four variables:
  - $Weather$: independent of the other variables

Bayesian Net



- A directed graph
- **Each node**: a R.V. (discrete or continuous)
- **Directed links:** connect pairs of nodes
  - A link from X to Y: X is a parent of Y
  - e.g. Cavity: a parent of Toothache

Bayesian Net



- No directed cycles
  - A directed acyclic graph (**DAG**)
  - Equivalent to say: there exists an ordering of the nodes such that links always go from lower numbered nodes to higher numbered nodes
- Conditional probability

  $P(X_i \,|\, Parents(X_i))$

  - Quantifies the effect of the parents on the child node

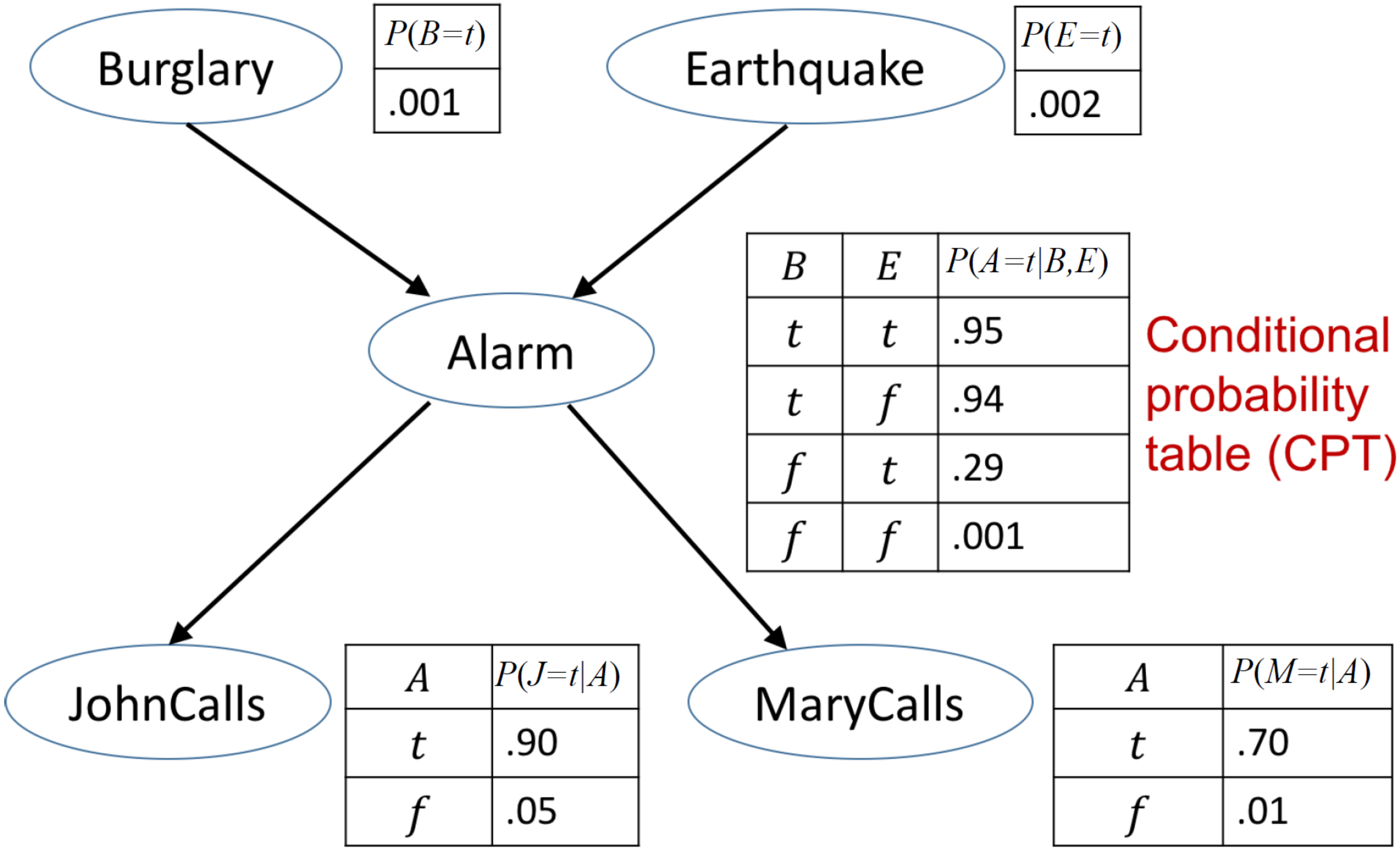
Burglar Alarm Example

- You have a new burglar alarm installed at home
- The alarm also responds on occasion to minor earthquakes
- You have two neighbors John and Mary
  - They promised to call you at work when they hear the alarm

Burglar Alarm Example

- John
  - Nearly always calls you when he hears the alarm
  - but sometimes confuses the telephone ringing with the alarm and calls you too
- Mary
  - Likes loud music and often misses the alarm
- Given the evidence of who has or has not called you, we would like to **estimate the probability of a burglary**

Bayesian Net



$P(B=t)$
.001

$P(E=t)$
.002

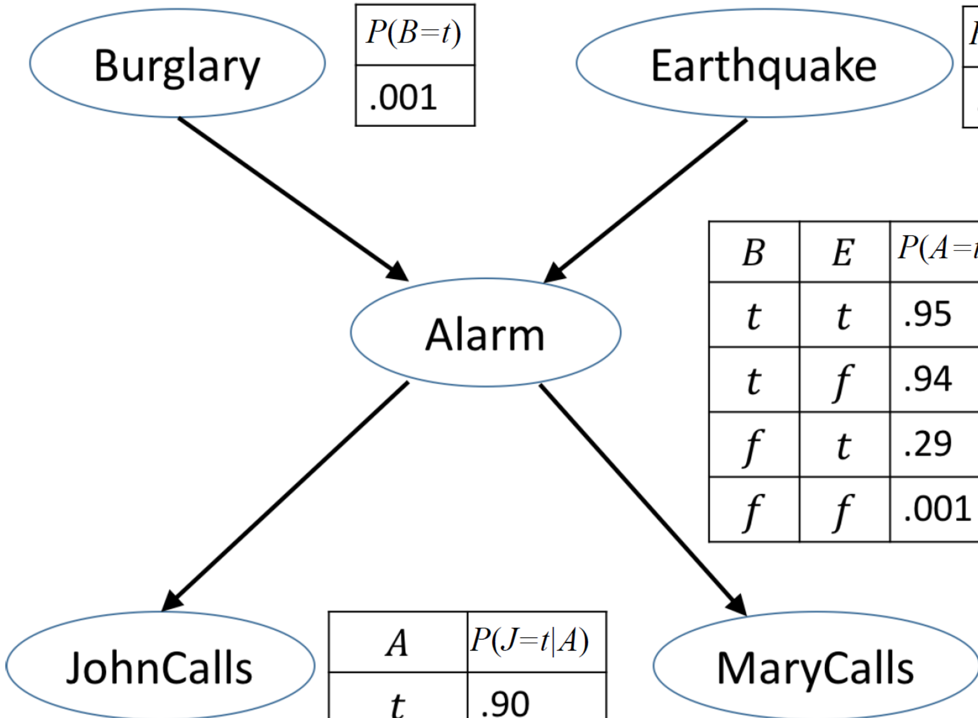
B	E	$P(A=t \mid B,E)$
t	t	.95
t	f	.94
f	t	.29
f	f	.001

A	$P(J=t \mid A)$
t	.90
f	.05

A	$P(M=t \mid A)$
t	.70
f	.01

- Burglary and Earthquake directly affect the Alarm

Bayesian Net



$P(B=t)$
.001

$P(E=t)$
.002

B	E	$P(A=t\|B,E)$
t	t	.95
t	f	.94
f	t	.29
f	f	.001

Conditional probability table (CPT)

A	$P(J=t\|A)$
t	.90
f	.05

A	$P(M=t\|A)$
t	.70
f	.01

- Whether John and Mary call depends only on the alarm
  - They do not perceive burglaries directly
  - They do not notice minor earthquakes
  - They do not confer before calling

Core Problems

- How to construct a Bayes Net?

- How to design inference procedures?
- Inference
  - e.g. given the evidence that John called, infer whether the Burglary occurred or not.

  - Find the probability…

Joint Distribution

- Chain rule (joint distribution of n variables)
- $P(X_1, \ldots, X_n) = P(X_n | X_{n-1}, \ldots, X_1)$
  $\times \mathrm{P}(X_{n-1} | X_{n-2}, \ldots, X_1)$
  $\times \cdots \times P(X_2 | X_1) \times P(X_1)$
  $= \prod_{i=1}^{n} P(X_i | X_{i-1}, \ldots, X_1)$

- Joint distribution of all variables in the **Bayes Net**:
  - The product of the conditional probabilities of all variables
  - $P(X_1, \ldots, X_n) = \prod_{i=1}^{n} P(X_i | parents(X_i))$

Construct Bayesian Networks

- Nodes
  - Determine the set of variables that are required to model the problem

- Links
  - For each node (variable), figure out its parent node(s), if there is any

    e.g. P(MaryCalls|Alarm)

  - Draw a link pointing from the parent node to the child node

Construct Bayesian Networks

- Number the nodes in a proper way
  - Parents have smaller order indices than their children
- $parents(X_i) \subseteq \{X_{i-1}, \ldots, X_1\}$
- $P(X_i | X_{i-1}, \ldots, X_1) = P(X_i | parents(X_i))$

- Joint distribution of all variables in the **Bayes Net**:
  - The product of the conditional probabilities of all variables
  - $P(X_1, \ldots, X_n) = \prod_{i=1}^{n} P(X_i | parents(X_i))$

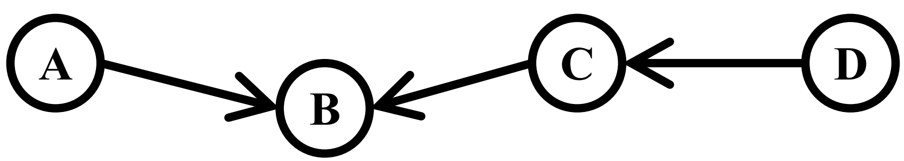
Construct Bayesian Networks

- CPTs: Write out the conditional probability table for each variable that has parent node(s)
  - $P(X_i|parents(X_i))$

Example

- Draw the Bayesian Network that corresponds to the factored joint probability distribution
- $P(A,B,C,D)= P(A)\times P(B|A, C) \times P(C|D) \times P(D)$

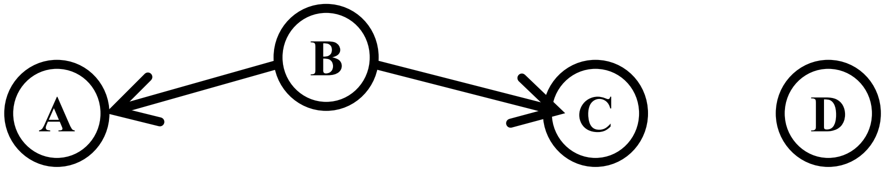
- Answer



Example

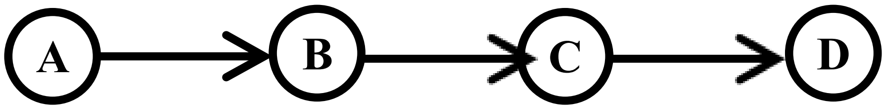
- Draw the Bayesian Network that corresponds to the factored joint probability
- $P(A,B,C,D) = P(A|B) \times P(B) \times P(C|B) \times P(D)$

- Answer



Example

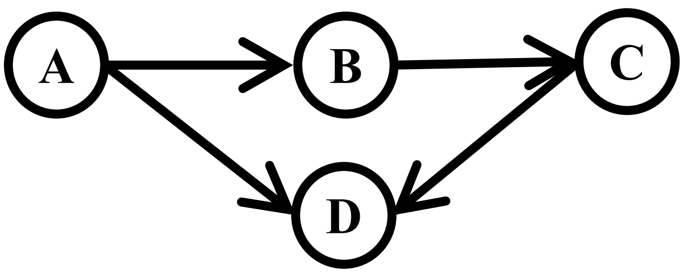
- Write the joint probability P(A,B,C,D) in terms of the product of conditional probabilities (factored joint probability) for the following Bayesian Network



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

Example

- Write the joint probability P(A,B,C,D) in terms of the product of conditional probabilities for the following Bayesian Network



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

©Ying Liu, Santa Clara University

Exact Inference in Bayesian Networks

- Task: to compute the posterior probability distribution for a set of **query variables**, given some observed **event** (some assignment of values to a set of **evidence variables**)

Exact Inference in Bayesian Networks

- A simple model: consider only one query variable
  - Query variable: X
  - Evidence variables: $\mathbf{E} = \{ E_1, \dots, E_m\}$
  - The observed event: $\mathbf{e} = \{E_1 = e_1, \dots, E_m = e_m\}$
  - Hidden variables:

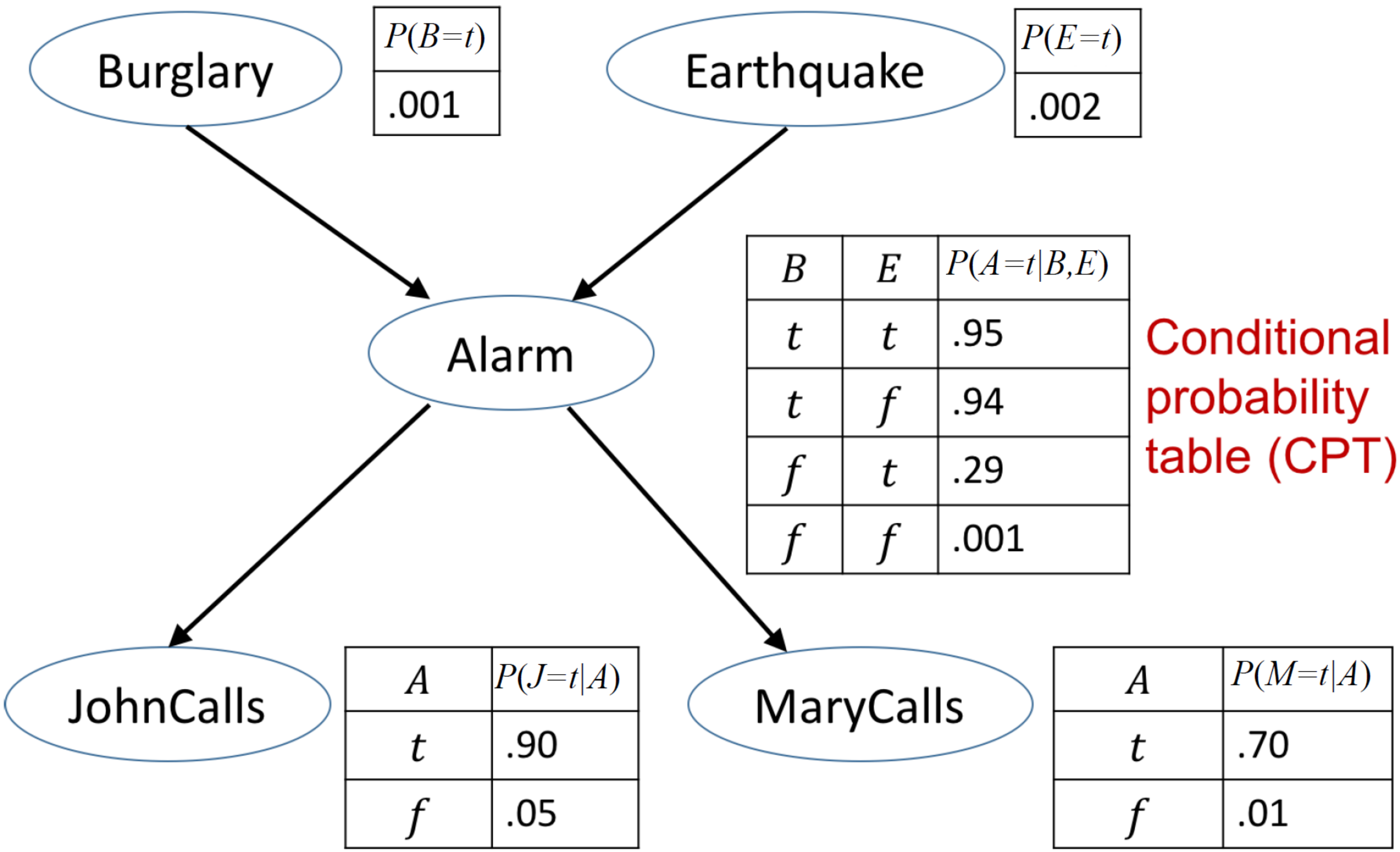
    Nonevidence, nonquery variables

    $\mathbf{Y} = \{Y_1, \dots, Y_l\}$

  - The complete set of variables: $\mathbf{V} = \{X\} \cup \mathbf{E} \cup \mathbf{Y}$

  - The query: ask for the posteriori probability distribution $\mathrm{P}(X|\mathbf{e})$

Burglary Alarm



$P(B=t)$
.001

$P(E=t)$
.002

B	E	$P(A=t\|B,E)$
t	t	.95
t	f	.94
f	t	.29
f	f	.001

A	$P(J=t\|A)$
t	.90
f	.05

A	$P(M=t\|A)$
t	.70
f	.01

Burglary Alarm

- Query variable: $B\ (Burglary)$

- B can take two values:

  B=b (Burglary occurs)

  B=$\neg b$ (Burglary does not occur)

Exact Inference in Bayesian Networks

- A simple model
  - Evidence variables:
  - $J\ (JohnCalls)$
    - $J = j$: John called
    - $J = \neg j$: John didn't call
  - $M(MaryCalls)$

  - The observed event:
  - $J = j, M = m$
  - That is, both John and Mary called

  - If $J = j, M = \neg m$, this means John called but Mary didn't call

Burglary Alarm

- A simple model
  - Hidden variables:

    Nonevidence, nonquery variables
  - E $(Earthquake)$
    - $E = e$, or $E = \neg e$
  - A $(Alarm)$
    - $A = a$, or $A = \neg a$
  - The query: ask for the posteriori probability distribution $P(B|\, j, m)$
    - That is, find $P(B = b|\, j, m)$ and $\mathrm{P}(B = \neg b|\, j, m)$
  - Which variables are not included in $P(B|\, j, m)$?
  - E, A

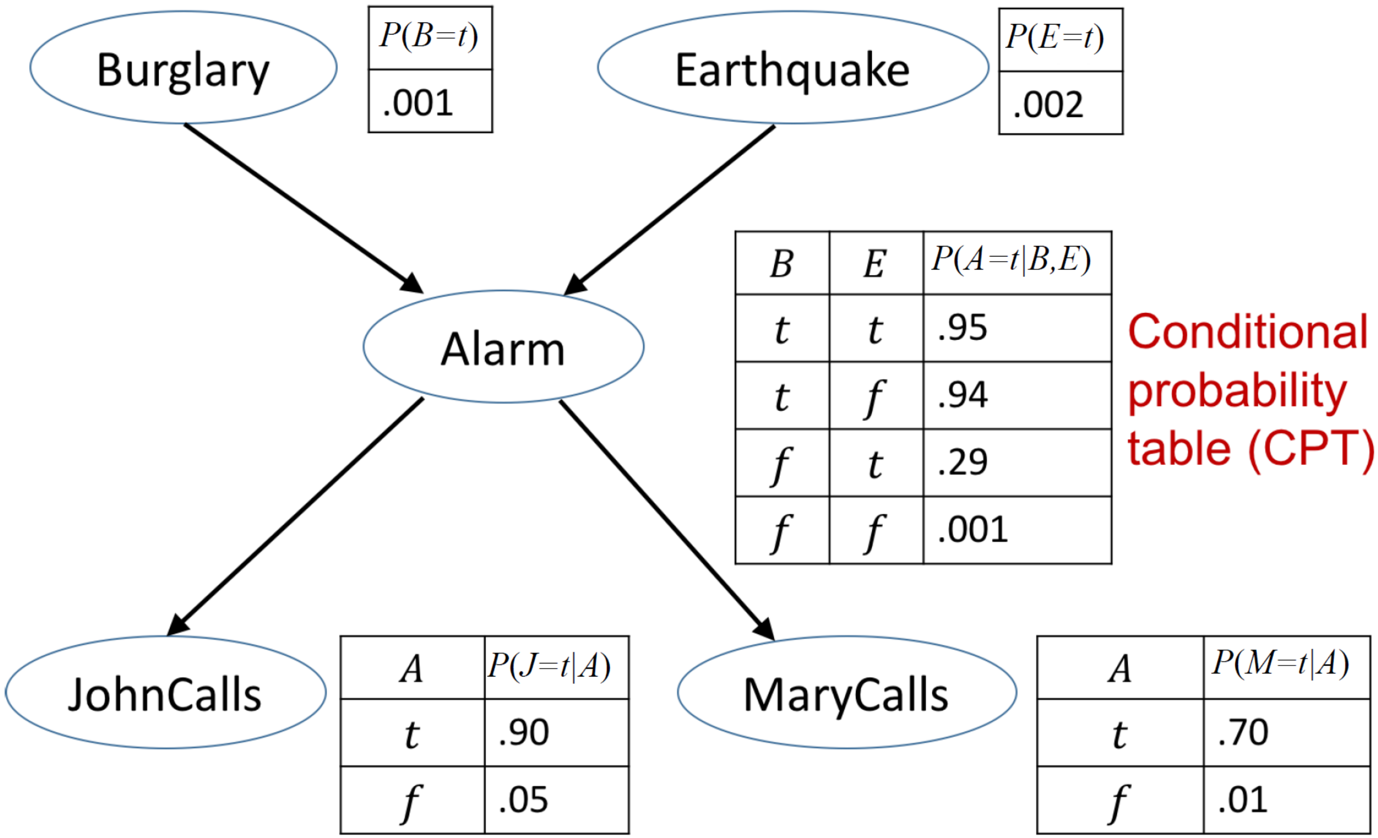
Joint Distribution

- Joint distribution of all variables in the **Bayes Net**:
  - The product of the conditional probabilities of all variables
  - $P(X_1, \ldots, X_n) = \prod_{i=1}^{n} P(X_i | parents(X_i))$

- Example: Calculate the joint probability that
  - The alarm has sounded
  - Neither a burglary nor an earthquake has occurred
  - Both John and Mary called
  - That is: $P(j, m, a, \neg b, \neg e)$

©Ying Liu, Santa Clara University

Joint Distribution



$P(B=t)$
.001

$P(E=t)$
.002

B	E	$P(A=t\|B,E)$
t	t	.95
t	f	.94
f	t	.29
f	f	.001

A	$P(J=t\|A)$
t	.90
f	.05

A	$P(M=t\|A)$
t	.70
f	.01

- $P(j, m, a, \neg b, \neg e)$

$$= P(j|a)P(m|a)P(a|\neg b \wedge \neg e)P(\neg b)P(\neg e)$$
$$= 0.90 \times 0.70 \times 0.001 \times 0.999 \times 0.998 = 0.000628$$

Reduce variable(s) in a joint distribution

- Joint distribution $P(X, E, Y)$

	$X = 0, E = 0$	$X = 0, E = 1$	$X = 1, E = 0$	$X = 1, E = 1$
$Y = 0$	0.1	0.2	0.05	0.05
$Y = 1$	0.3	0.1	0.1	0.1

- $P(X = 0, E = 0, Y = 1) =$0.3
- $P(X = 1, E = 0, Y = 0) = 0.05$
- …

Reduce variable(s) in a joint distribution

- Joint distribution $P(X, E, Y)$

	$X = 0, E = 0$	$X = 0, E = 1$	$X = 1, E = 0$	$X = 1, E = 1$
$Y = 0$	0.1	0.2	0.05	0.05
$Y = 1$	0.3	0.1	0.1	0.1

- $P(X, E) = \sum_{Y=0,1} P(X, E, Y)$
- $P(X = 0, E = 0)$

$$= \sum_{Y=0,1} P(X = 0, E = 0, Y)$$
$$= P(X = 0, E = 0, Y = 0)$$
$$+ P(X = 0, E = 0, Y = 1)$$
$$= 0.1 + 0.3 = 0.4$$

Reduce variable(s) in a joint distribution

- Joint distribution $P(X, E, Y)$

	$X = 0, E = 0$	$X = 0, E = 1$	$X = 1, E = 0$	$X = 1, E = 1$
$Y = 0$	0.1	0.2	0.05	0.05
$Y = 1$	0.3	0.1	0.1	0.1

- $P(X, E) = \sum_{Y=0,1} P(X, E, Y)$
- $P(X = 1, E = 0)$

$$= \sum_{Y=0,1} P(X = 1, E = 0, Y)$$
$$= P(X = 1, E = 0, Y = 0)$$
$$+ P(X = 1, E = 0, Y = 1)$$
$$= 0.05 + 0.1 = 0.15$$

Reduce variable(s) in a joint distribution

- Joint distribution $P(X, E, Y)$

	$X = 0, E = 0$	$X = 0, E = 1$	$X = 1, E = 0$	$X = 1, E = 1$
$Y = 0$	0.1	0.2	0.05	0.05
$Y = 1$	0.3	0.1	0.1	0.1

- $P(Y) = \sum_{X=0,1} \sum_{E=0,1} P(X, E, Y)$
- $P(Y = 0)$

$$= \sum_{X=0,1} \sum_{E=0,1} P(X, E, Y = 0)$$

$$= P(X = 0, E = 0, Y = 0) + P(X = 0, E = 1, Y = 0) + P(X = 1, E = 0, Y = 0) + P(X = 1, E = 1, Y = 0)$$

$$= 0.1 + 0.2 + 0.05 + 0.05 = 0.4$$

Reduce variable(s) in a joint distribution

- Joint distribution $P(X, E, Y)$

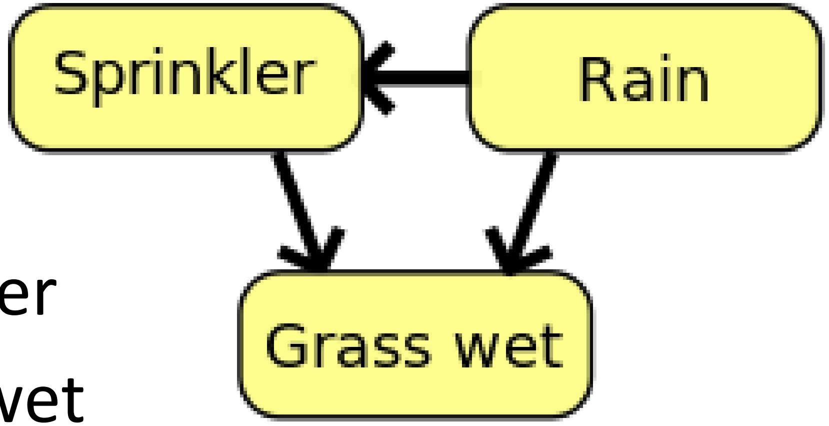
	$X = 0, E = 0$	$X = 0, E = 1$	$X = 1, E = 0$	$X = 1, E = 1$
$Y = 0$	0.1	0.2	0.05	0.05
$Y = 1$	0.3	0.1	0.1	0.1

- $P(Y) = \sum_{X=0,1;E=0,1} P(X, E, Y)$
- $P(Y = 0) = \sum_{X=0,1;E=0,1} P(X, E, Y = 0)$

$$= P(X = 0, E = 0, Y = 0) + P(X = 0, E = 1, Y = 0) + P(X = 1, E = 0, Y = 0) + P(X = 1, E = 1, Y = 0)$$

$$= 0.1 + 0.2 + 0.05 + 0.05 = 0.4$$

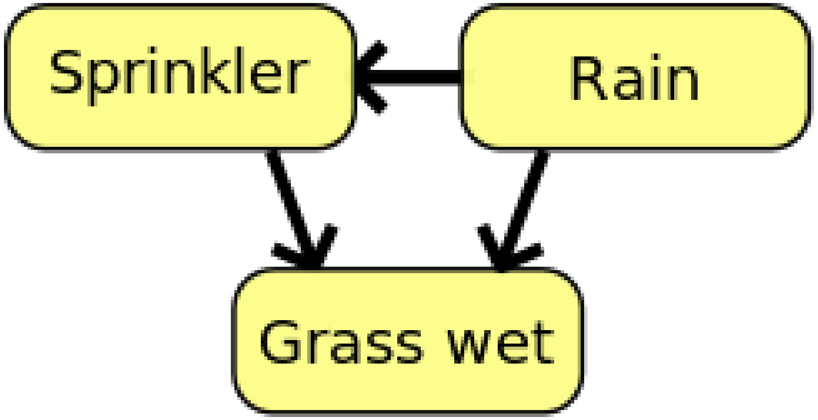
Example



- Rain is the cause of Sprinkler
- Rain is the cause of Grass wet
- Sprinkler is the cause of Grass wet
- Goal: find the distribution of
- $P(Rain|Grass\ wet\ = t)$
- That is, find
  - $P(Rain = t|Grass\ wet\ = t)$
  - $P(Rain = f|Grass\ wet\ = t)$
  - They sum up to 1

©Ying Liu, Santa Clara University

Example



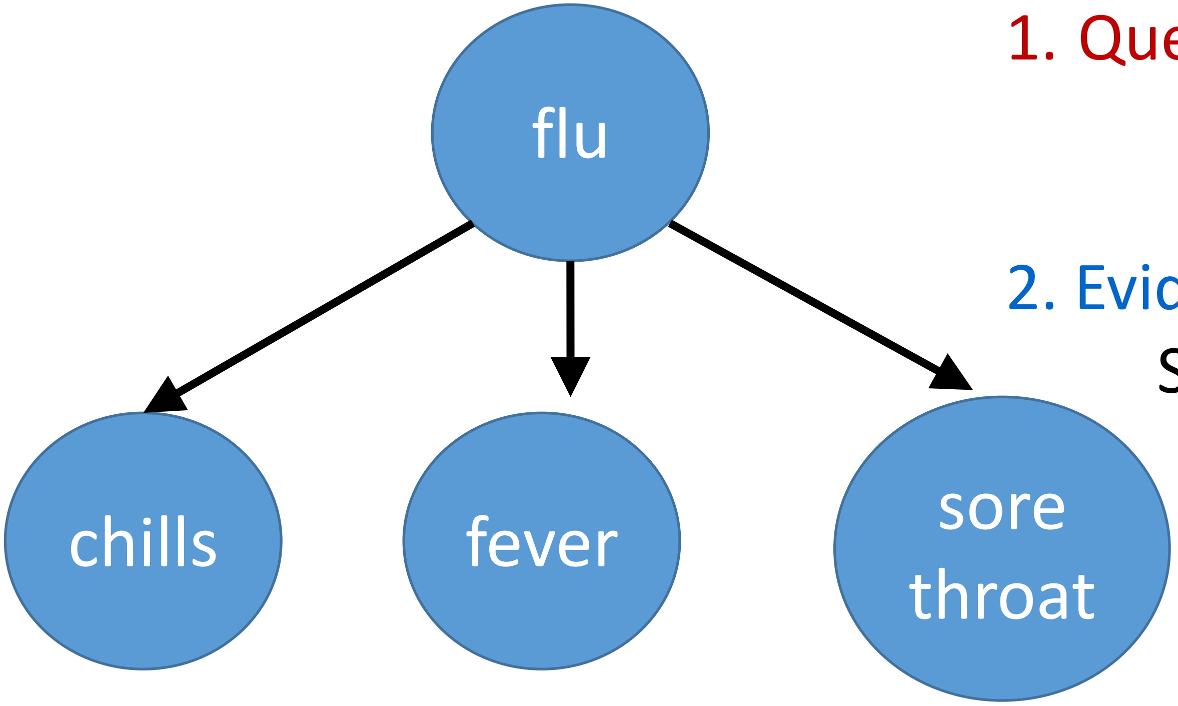
1. Query variable?
   Rain: R

2. Evidence variable?
   Grass wet: G

3. Hidden variable?
   Sprinkler: S

- R: Rain
- G: Grass
- S: Sprinkler
- Goal: find $P(R = t|G = t)$ and $P(R = f|G = t)$
- Which variable is not included in these two probabilities?

Disease and its symptoms



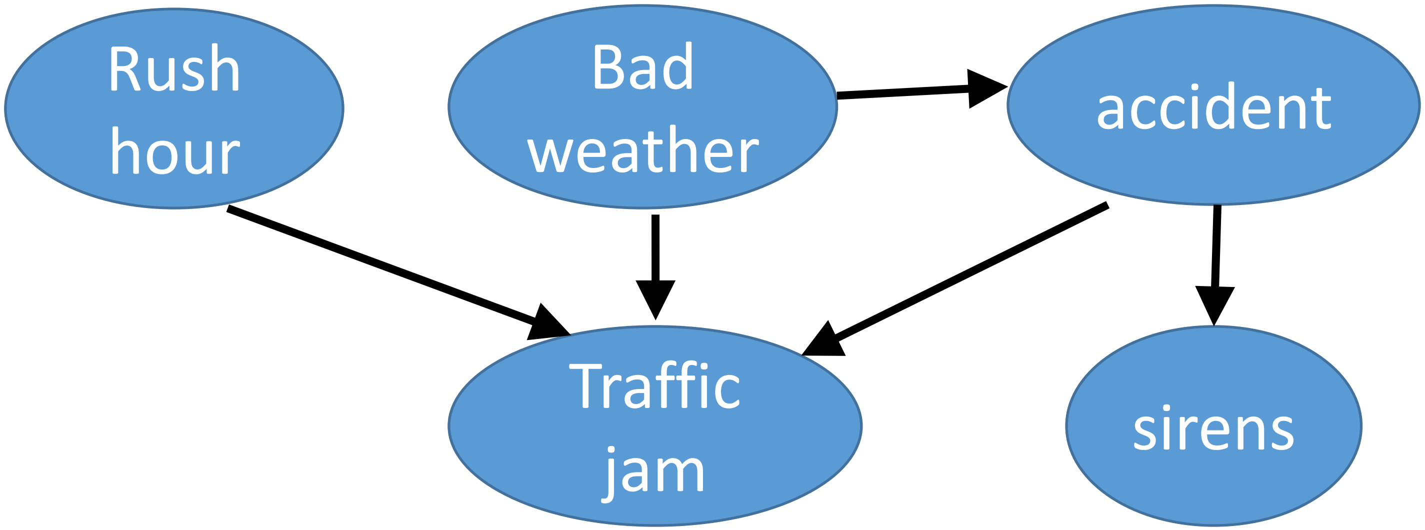
1. Query variable?
flu

2. Evidence variable?
Sore throat

- Find P(flu|sore throat)

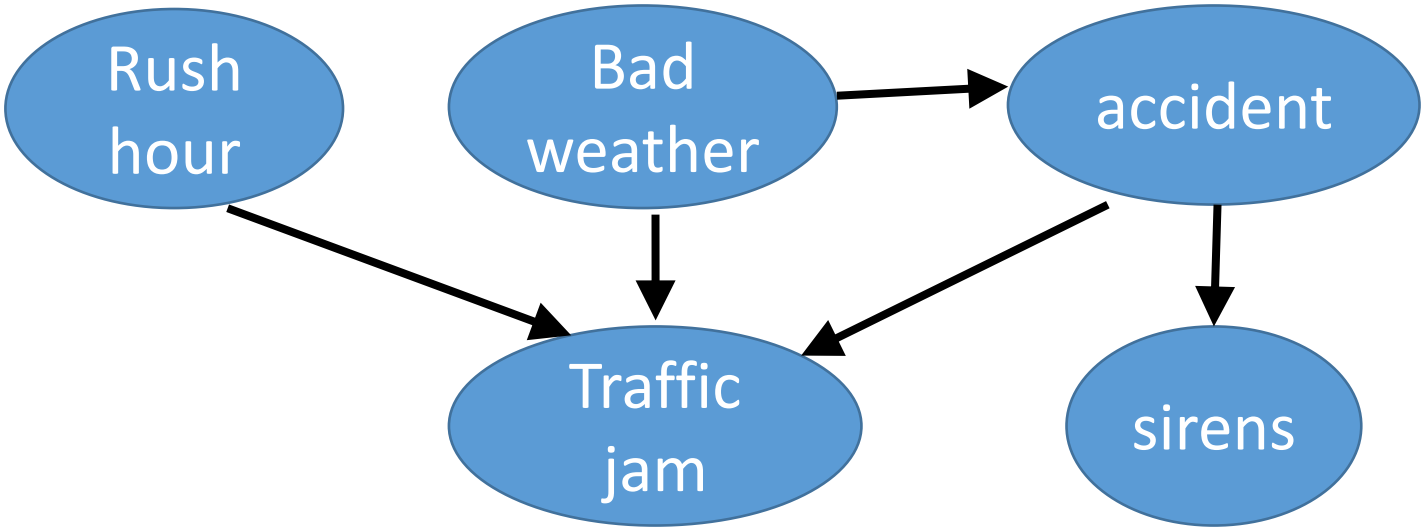
3. Hidden variable(s)?
Chills, fever

Events surrounding a traffic jam



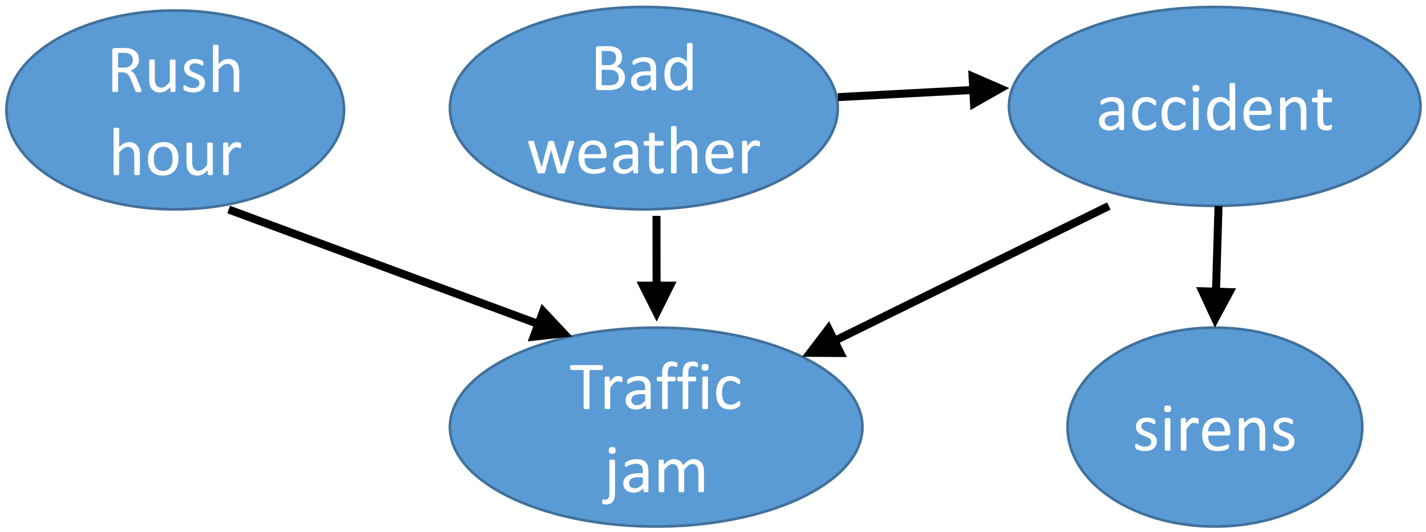
- Find P(bad weather|traffic jam, sirens)
- 1. Query variable?
- Bad weather

Events surrounding a traffic jam



- Find P(bad weather|traffic jam, sirens)
- 2. Evidence variable(s)?
- Traffic jam, sirens

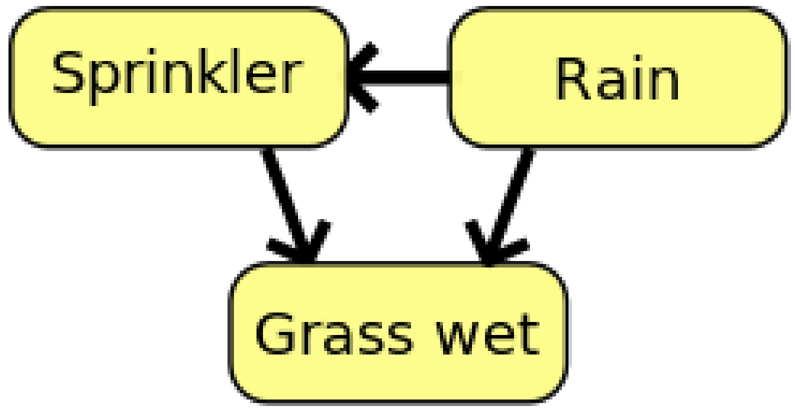
Events surrounding a traffic jam



- Find P(bad weather|traffic jam, sirens)
- 3. Hidden variable(s)?
- Rush hour, accident

Inference

P(S|R)	**Sprinkler**	
Rain	t	f
f	0.4	0.6
t	0.01	0.99



Rain	
t	f
0.2	0.8

P(G|S,R)		**Grass Wet**	
Sprinkler	Rain	t	f
f	f	0.0	1.0
f	t	0.8	0.2
t	f	0.9	0.1
t	t	0.99	0.01

Inference:
$P(R|G = t)$

Inference

- $P(R = t|G = t) = \frac{P(R=t,G=t)}{P(G=t)}$ (1)
- $P(R = f|G = t) = \frac{P(R=f,G=t)}{P(G=t)}$ (2)
- $P(G = t)$ is the same in (1) and (2)

- $Let\ \alpha = \frac{1}{P(G=t)}, then$
- $P(R = t|G = t) = \alpha P(R = t, G = t)$
- Which variable is not utilized?
- S!

©Ying Liu, Santa Clara University

Inference

- $P(R = t|G = t) = \frac{P(R=t,G=t)}{P(G=t)}$ (1)
- $P(R = f|G = t) = \frac{P(R=f,G=t)}{P(G=t)}$ (2)
- $P(G = t)$ is the same in (1) and (2)

- $Let\ \alpha = \frac{1}{P(G=t)}, then$
- $P(R = t|G = t) = \alpha P(R = t, G = t)$

  $= \alpha \sum_{S=t,f} P(R = t, G = t, S)$
- Sum over hidden variable S

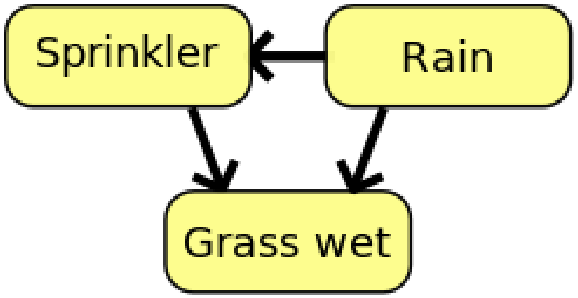
Inference

- $P(R=t|G=t) = \alpha P(R=t, G=t)$
- $= \alpha \sum_{S=t,f} P(R=t, G=t, S)$ (*)
- Recall

$$P(X_1, \ldots, X_n) = \prod_{i=1}^{n} P(X_i | parents(X_i))$$

- (*)$= \alpha \sum_{S=t,f} P(R=t) P(S|R=t) P(G=t|R=t,S)$

$$= \alpha \underbrace{P(R=t)}_{0.2} \sum_{S=t,f} P(S|R=t) P(G=t|R=t,S)$$



Rain	
t	f
0.2	0.8

Inference

- $= \alpha \underbrace{P(R=t)}_{0.2} \sum_{S=t,f} P(S|R=t)P(G=t|R=t,S)$
- $= \alpha \times 0.2 \times [\underbrace{P(S=t|R=t)}_{0.01} \times \underbrace{P(G=t|R=t,S=t)}_{0.99}$

  $+ \underbrace{P(S=f|R=t)}_{0.99} \times \underbrace{P(G=t|R=t,S=f)}_{0.8}]$
- $= \alpha \times 0.1604$

P(S|R)	**Sprinkler**	
Rain	t	f
f	0.4	0.6
t	0.01	0.99

P(G|S,R)		**Grass Wet**	
Sprinkler	Rain	t	f
f	f	0.0	1.0
f	t	0.8	0.2
t	f	0.9	0.1
t	t	0.99	0.01

Inference

- $P(R = t|G = t) = \frac{P(R=t,G=t)}{P(G=t)}$ (1) $= \alpha \times 0.1604$
- $P(R = f|G = t) = \frac{P(R=f,G=t)}{P(G=t)}$ (2)
- $P(G = t)$ is the same in (1) and (2)

- $Let\ \alpha = \frac{1}{P(G=t)}, then$
- $P(R = f|G = t) = \alpha P(R = f, G = t)$
  $= \alpha \sum_{S=t,f} P(R = f, G = t, S)$
- Sum over hidden variable S

Inference

- $P(R=f|G=t) = \alpha P(R=f, G=t)$
- $= \alpha \sum_{S=t,f} P(R=f, G=t, S)$
- $= \alpha \sum_{S=t,f} P(R=f) P(S|R=f) P(G=t|R=f, S)$

- $= \alpha \underbrace{P(R=f)}_{0.8} \sum_{S=t,f} P(S|R=f) P(G=t|R=f, S)$

Rain	
t	f
0.2	0.8

Inference

- $= \alpha \underbrace{P(R=f)}_{0.8} \sum_{S=t,f} P(S|R=f)P(G=t|R=f,S)$
- $= \alpha \times 0.8 \times [\underbrace{P(S=t|R=f)}_{0.4} \times \underbrace{P(G=t|R=f,S=t)}_{0.9}$
  $+ \underbrace{P(S=f|R=f)}_{0.6} \times \underbrace{P(G=t|R=f,S=f)}_{0}]$
- $= \alpha \times 0.288$

P(S|R)	**Sprinkler**	
Rain	t	f
f	0.4	0.6
t	0.01	0.99

P(G|S,R)		**Grass Wet**	
Sprinkler	Rain	t	f
f	f	0.0	1.0
f	t	0.8	0.2
t	f	0.9	0.1
t	t	0.99	0.01

Inference

- $P(R = t \,\big|\, G = t) = \alpha \times 0.1604$
- $P(R = f \,\big|\, G = t) = \alpha \times 0.288$

- $\alpha \times 0.1604 + \alpha \times 0.288 = 1, \alpha = 2.2302$

- Distribution:
- $P(R = t|G = t) = \alpha \times 0.1604 = 0.3577$

- $P(R = f|G = t) = \alpha \times 0.288 = 0.6423$

Inference

- $P(R = t \mid G = t) = \alpha \times 0.1604$
- $P(R = f \mid G = t) = \alpha \times 0.288$

- Distribution (another way to calculate it):

- $P(R = t|G = t) = \frac{\alpha\times0.1604}{\alpha\times0.1604+\alpha\times0.288} = 0.3577$

- $P(R = f|G = t) = \frac{\alpha\times0.288}{\alpha\times0.1604+\alpha\times0.288} = 0.6423$

©Ying Liu, Santa Clara University