COMP90051 **Statistical Machine Learning**

Semester 2, 2015

PGM (Bayesian Network)

THE UNIVERSITY OF MELBOURNE

Random Variables and Events

- *Random variables*: the (uncertain) state of the world
  - Denoted by capital letters
  - ***R***: *Is it raining?* (binary, discrete)
  - *S*: What's the wind speed? (continuous)
- *Atomic event*: a complete assignment of domain values to random variables
  - ***R*** = True
  - *S* = 100 km/h
  - Atomic events are mutually exclusive and exhaustive

Joint Probability Distributions

- If the world consists of only two Boolean variables A and B, then there are four distinct atomic events:
    - $A = True \wedge B = True$
    - $A = True \wedge B = False$
    - $A = False \wedge B = True$
    - $A = False \wedge B = False$
- A *joint probability distribution* is an assignment of probabilities to every possible atomic event

Atomic event	P
$A = True \wedge B = True$	0.1
$A = True \wedge B = False$	0.2
$A = False \wedge B = True$	0.3
$A = False \wedge B = False$	0.4

Joint Probability Distributions

- What's the size of the table given n variables with d domain values?
- Notation:
  - $P(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n)$ refers to a single entry in the joint probability distribution table;

Atomic event	P
$A = True \wedge B = True$	0.1

  - $P(X_1, X_2, \ldots, X_n)$ refers to the entire joint probability distribution table;

Marginalisation

- From the joint distribution $P(A, B)$ we can find the ***marginal distributions*** $P(A)$ and $P(B)$

Atomic event	P
$A = True \wedge B = True$	0.1
$A = True \wedge B = False$	0.2
$A = False \wedge B = True$	0.3
$A = False \wedge B = False$	0.4

Atomic event	P
$A = True$	0.3
$A = False$	0.7

Atomic event	P
$B = True$	0.4
$B = False$	0.6

Conditional Probability

- Definition:

$$P(A \mid B) = \frac{P(A,B)}{P(B)}$$

obtain the conditional probability with the joint

- Obtain the joint with the conditional probability:

$$P(A,B) = P(A \mid B)P(B) = P(B \mid A)P(A)$$

- The chain rule:

$$P(A_1,\ldots,A_n) = P(A_1)P(A_2 \mid A_1)P(A_3 \mid A_1,A_2)\ldots P(A_n \mid A_1,\ldots,A_{n-1})$$

$$= \prod_{i=1}^{n} P(A_i \mid A_1,\ldots,A_{i-1})$$

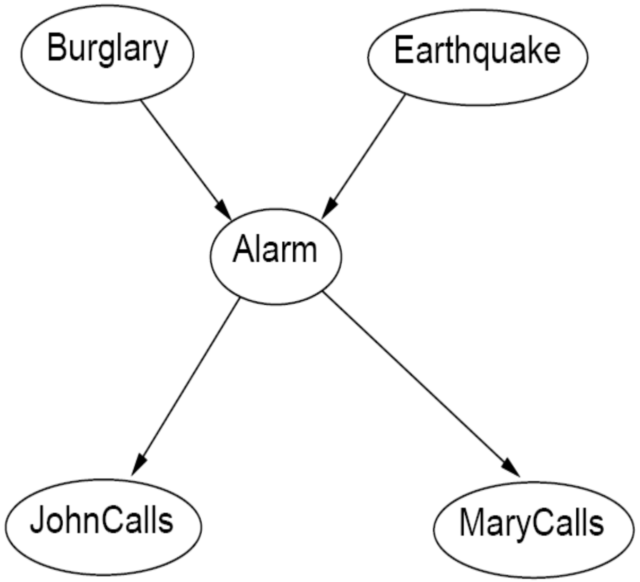
Independence

- A and B are *independent* iff:
  - $P(A \wedge B) = P(A) \cdot P(B)$, equivently:
  - $P(A|B) = P(A)$ and $P(B|A) = P(B)$
- *mutually exclusive* events ≠ independent events
  - For mutually exclusive: $P(A \vee B) = P(A) + P(B)$
- A and B are conditionally independent given C iff:
  - $P(A \wedge B|C) = P(A|C) \cdot P(B|C)$
  - E.g. naïve Bayesian:
    - $P(Y|\, X1, X2, X3) \propto P(X1|Y) \cdot P(X2|Y) \cdot P(X3|Y) \cdot P(Y)$

PGM: Bayesian Network

- A type of *graphical model*
- A Bayesian network is a Directed Acyclic Graph (DAG)
- A Bayesian network states conditional independence relationships between random variables
- Compact specification of full joint distributions

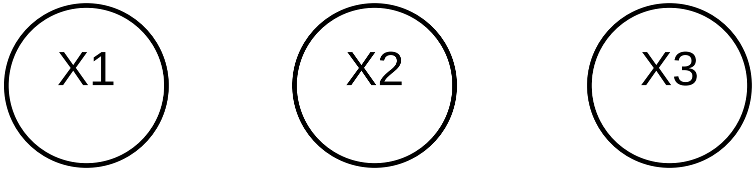
Bayesian Network



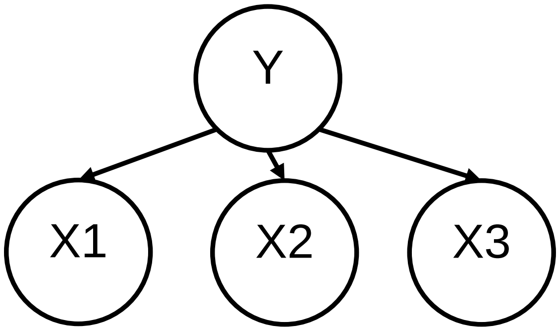
- Nodes: random variables
- Arcs: interactions
  - An arrow from one variable to another indicates direct influence
  - A node is conditionally independent of its non-descendants given its parent

Examples

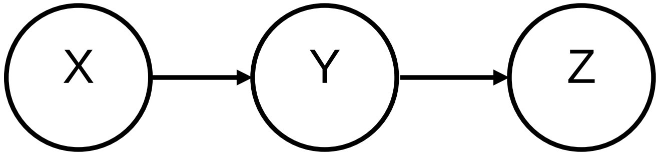
- Unconditionally/complete independent:



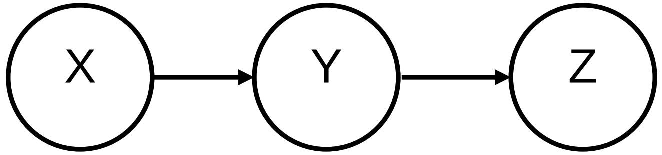
- Naïve Bayes: conditionally independent



- How about:

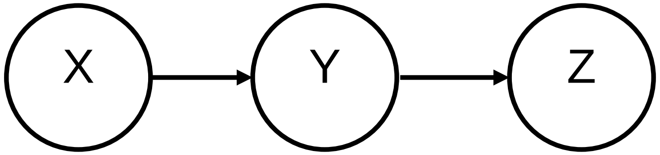


Examples

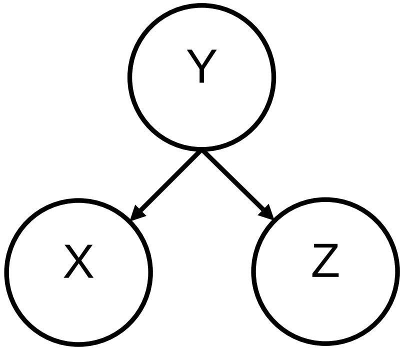


- Are X and Z independent?
- Is Z independent of X given Y?

Examples



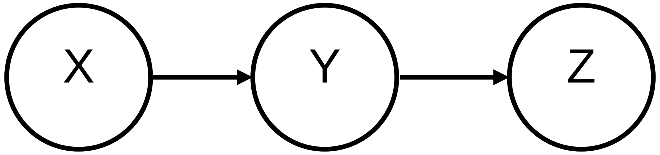
- Are X and Z independent?
- Is Z independent of X given Y?



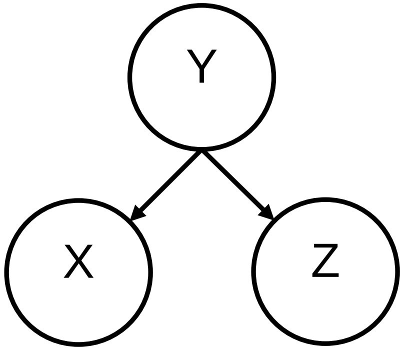
Common cause

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

Examples

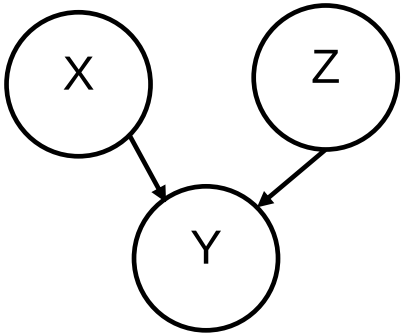


- Are X and Z independent?
- Is Z independent of X given Y?



Common cause

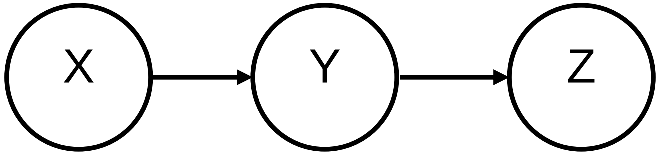
- Are X and Z independent?
- Are they conditionally independent given Y?



Common effect

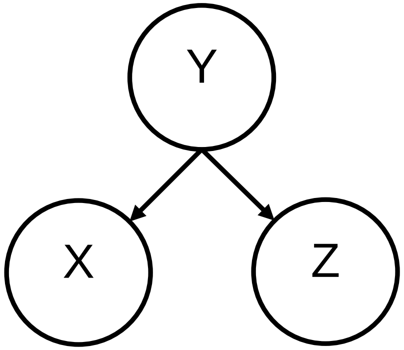
- Are X and Z independent?
- Are they conditionally independent given Y?

Examples



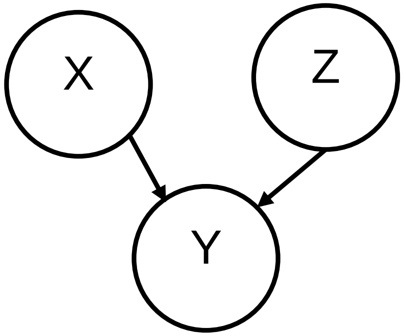
- Are X and Z independent? No
- Is Z independent of X given Y? Yes

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



Common cause

- Are X and Z independent? No
- Are they conditionally independent given Y? Yes

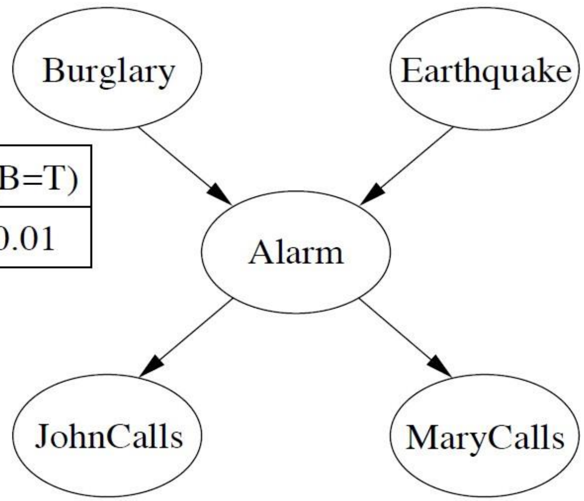


Common effect

- Are X and Z independent? Yes
- Are they conditionally independent given Y? No

PGM: Model Representation

- Directed acyclic graph
- Conditional probability table (parameters)



P(B=T)
0.01

P(E=T)
0.02

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

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

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

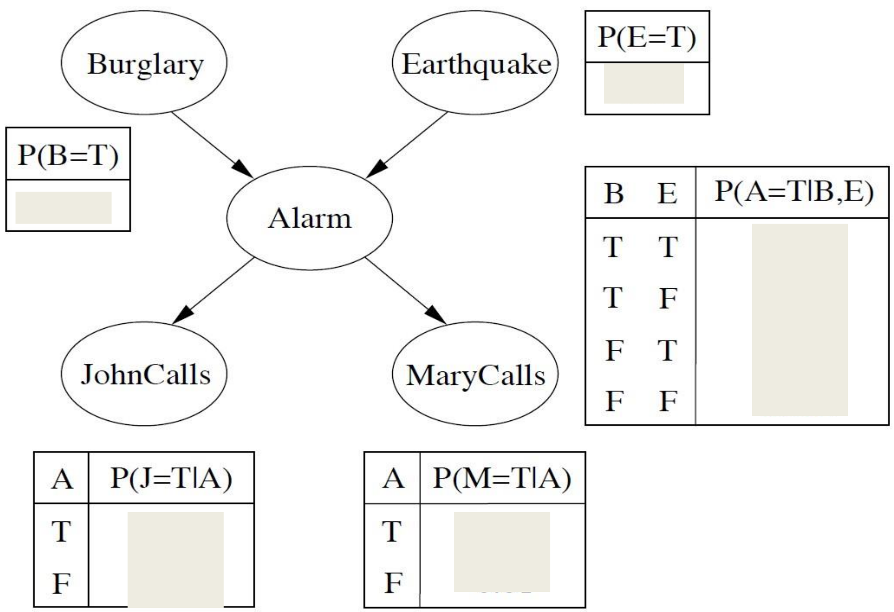
- Compact: just 10 rows vs 31 rows in a full joint table!

PGM: Training

- Constructing the structure of the network
  - domain expert to decide the causal relations
  - *structure learning* algorithms exist, but complicated

PGM: Training

- Constructing the structure of the network
  - domain expert to decide the causal relations
  - *structure learning* algorithms exist, but complicated
- Parameter learning (filling the table)



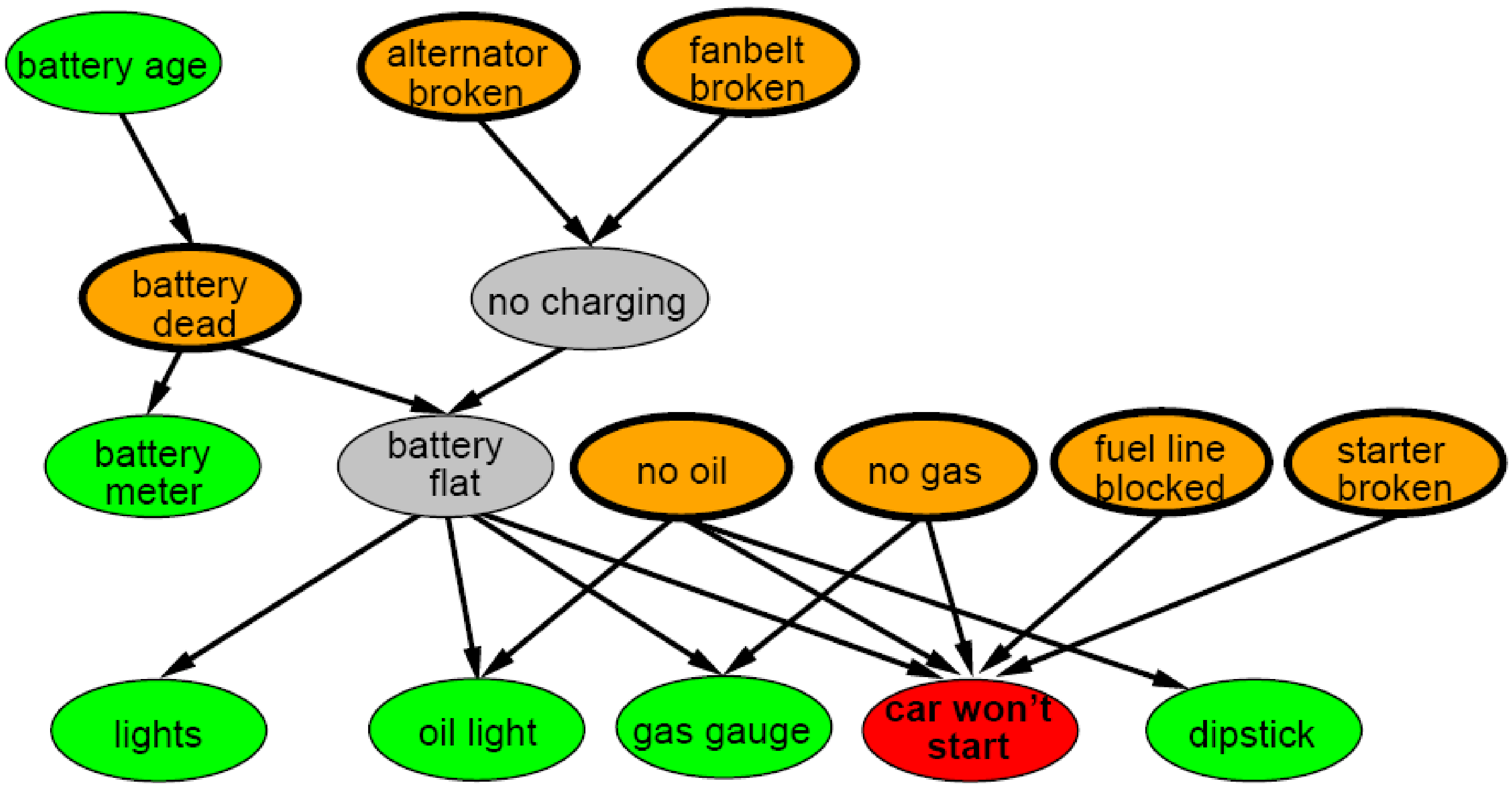
Training

A	B	E	J	M
T	F	T	F	T
F	T	F	F	F
T	F	T	F	T
...				

  - Using EM method if there are missing values

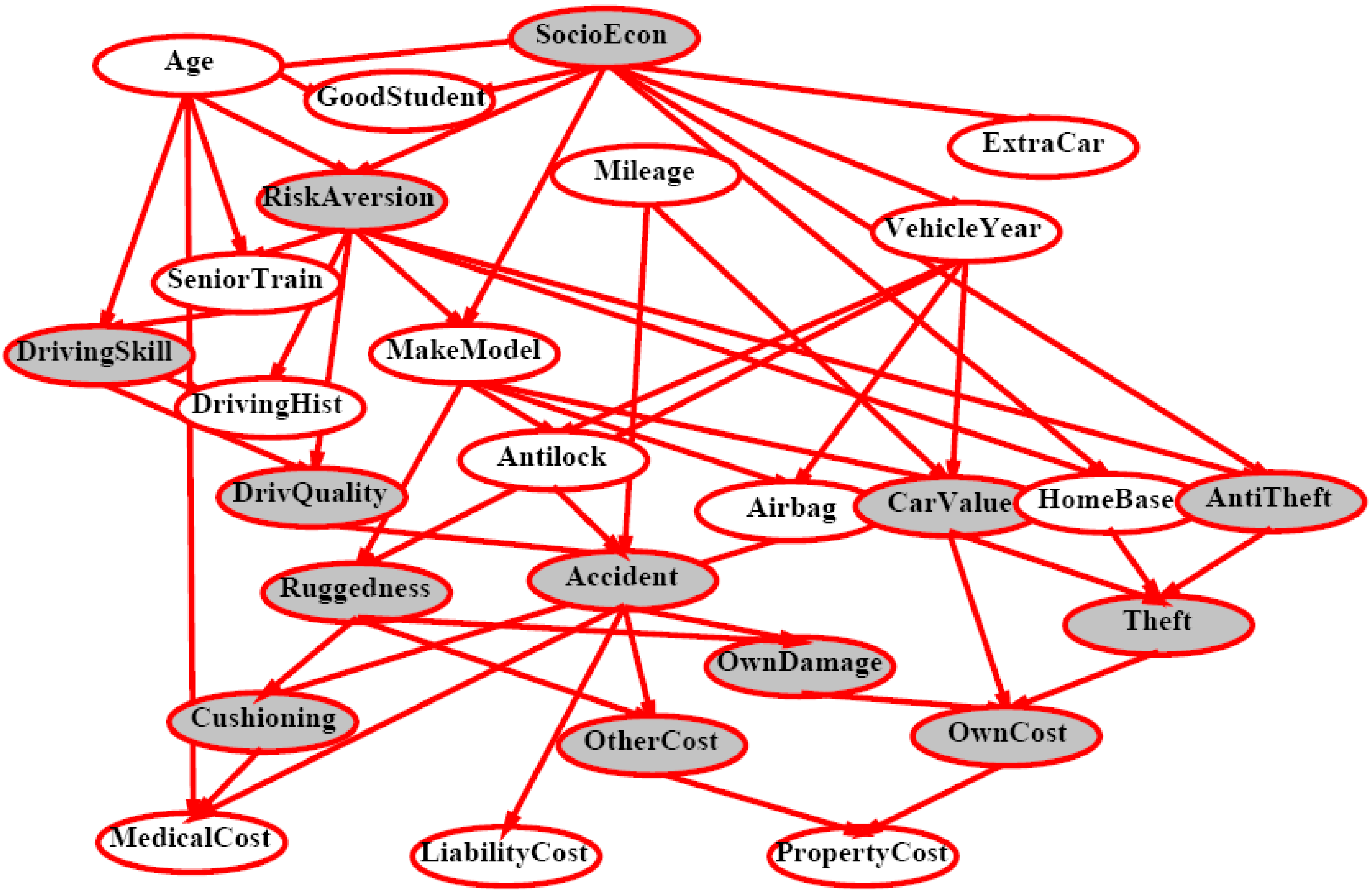
A more realistic Bayes Network: Car diagnosis

- Initial observation: car won't start
- Orange: "broken, so fix it" nodes
- Green: testable evidence
- Gray: "hidden variables" to ensure sparse structure, reduce parameteres



Source: UIUC Artificial Intelligence (CS440/ECE448)

Car insurance

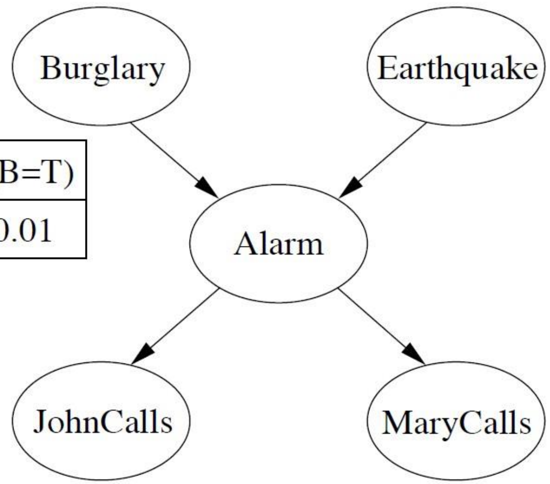


Source: UIUC Artificial Intelligence (CS440/ECE448)

PGM: Probability Inference

- A general scenario:
  - Query *variables:* $\boldsymbol{X}$
  - *Evidence* (*observed*) variables and their values: $\boldsymbol{E} = \boldsymbol{e}$
  - *Unobserved* variables: $\boldsymbol{Y}$
- Inference problem: answer questions about the query variables given the evidence variables
- Detail about probability inference will be explained next week .
  - Enumeration
  - Elimination Algorithm

Example: Probability Inference



P(B=T)
0.01

P(E=T)
0.02

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

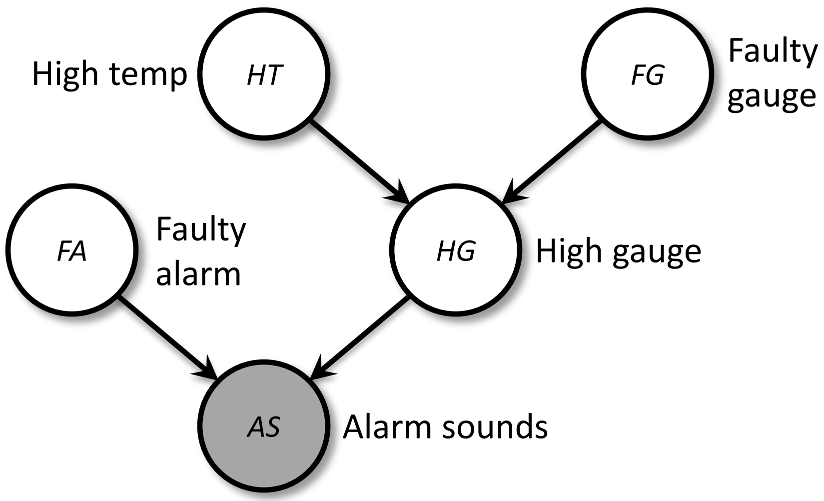
A	P(J=T|A)
T	0.90
F	0.05

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

Compute the probability that there is an earthquake given both John and Mary call.

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

Example: Probability Inference



Alarm sounds (evidence) meltdown? (query)

Reference

[1] Stuart Russell and Peter Norvig. Artificial Intelligence: A Modern Approach 3rd Edition.

[2] Kevin B. Korb and Ann E. Nicholson. Bayesian Artificial Intelligence 2nd Edition

[3] Some slides were derived from UIUC Artificial Intelligence (CS440/ECE448)