CS 520: Introduction to Artificial Intelligence

Prof. Louis Steinberg

Lecture 17: Bayes Nets

Acting under uncertainty

- **Probability: methods for representing and reasoning about uncertainty of knowledge**
- **Utility theory: a way to represent and reason about the *desirability* of different outcomes**
- **Decision theory: A way to combine probability and utility to decide what to do.**

Basic Properties

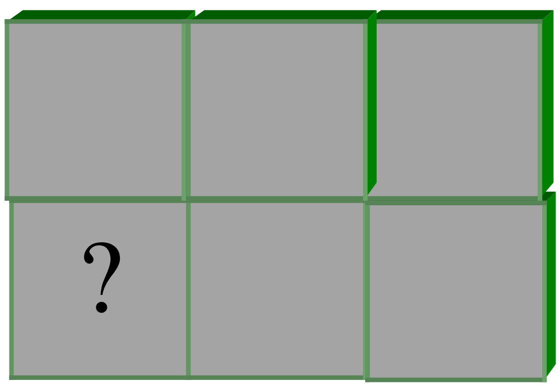
- $0 \leq P(A) \leq 1$
- $P(true) = 1$
- $P(false) = 0$
- $P(A \quad B) = P(A) + P(B) - P(A \ \square\ B)$
- $P(A \mid B) = P(A \ \square\ B) / P(B)$
  Equivalently: $P(A \wedge B) = P(A \mid B) * P(B)$

Example Use: Mine Sweeper

- **Grid of cells, some have mines**
- **Probe a cell:**
  - **If has a mine, lose**
  - **Else tells number of neighboring mines**
- **Variables:**
  - **M(x, y): boolean: mine**
  - **C(x, y): neighbor mine count**
  - **N: total mines**

Mine Sweeper

N = 3



P(M(1, 1)) =

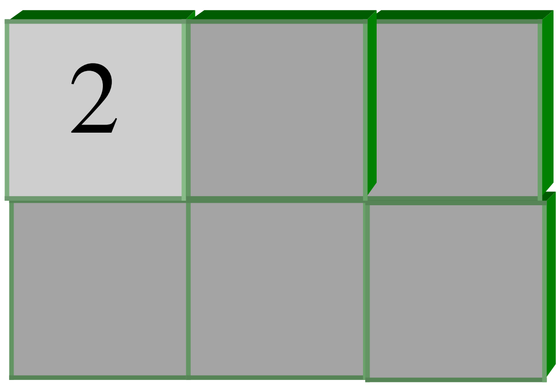
$$\frac{\text{\# ways of choosing 2 of 5}}{\text{\# ways of choosing 3 of 6}}$$

$$= \frac{5!/(2!3!)}{6!/(3!3!)}$$

$$= 10/20 = .5$$

Mine Sweeper

N = 3



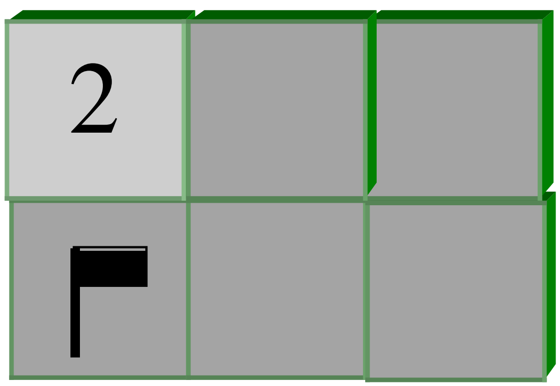
P(C(1,2)=2 ^ ~M(1,2)) =

$$\frac{\text{choose}(2, 3) * \text{choose}(1,2)}{\text{choose}(3, 6)}$$

= 3*2/20 = .3

Mine Sweeper

N = 3



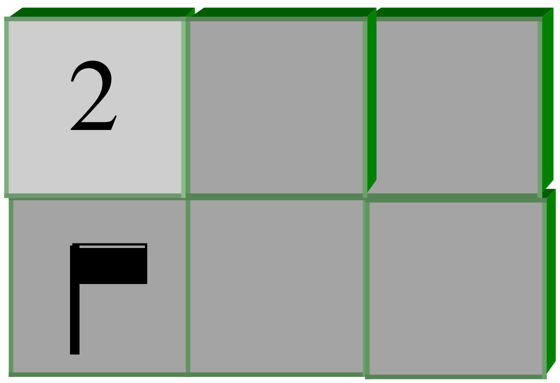
P(C(1,2)=2 ^ ~M(1,2) ^ M(1,1))=

$$\frac{\text{choose}(1, 2) * \text{choose}(1,2)}{\text{choose}(3, 6)}$$

= 2 * 2 / 20 = .2

Mine Sweeper

N = 3

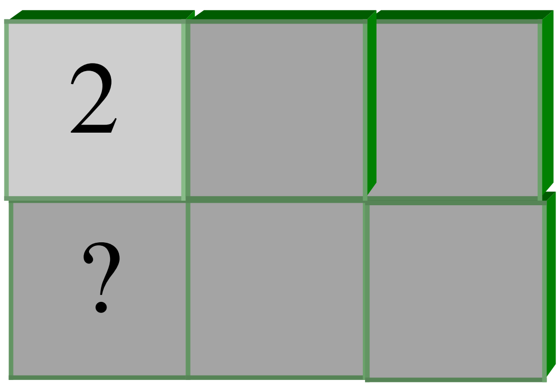


$$P(\, C(1,2)=2 \wedge \sim M(1,2) \mid M(1,1)\,) = \frac{P(\, C(1,2)=2 \wedge \sim M(1,2) \wedge M(1,1)\,)}{P(\, M(1,1)\,)}$$

$= .2/.5 = 2/5$

Mine Sweeper

N = 3



P(M(1,1) | C(1,2)=2 ^ ~M(1,2))=

P(C(1,2)=2 ^ ~M(1,2) | M(1,1))

*** P(M(1,1))**

/ P(C(1,2)=2 ^ ~M(1,2))

= 2/5 * 1/2 / 3/10

= 2/3

P(M(1,1) | C(1,2)=2 ^ ~M(1,2))=

choose(1,2)*choose(1,2)

choose(2,3)*choose(1,2)

= 4 / 6 = 2/3

Mine Sweeper

- **Known:**

  $P(M(1, 1) \vee M(2, 1)) = 1$

  $P(M(1, 1) \wedge M(2, 1)) = 0$

  $P(M(1, 1)) = P(M(2, 1))$

- **Conclude:**

  $1 = P(M(1, 1))$

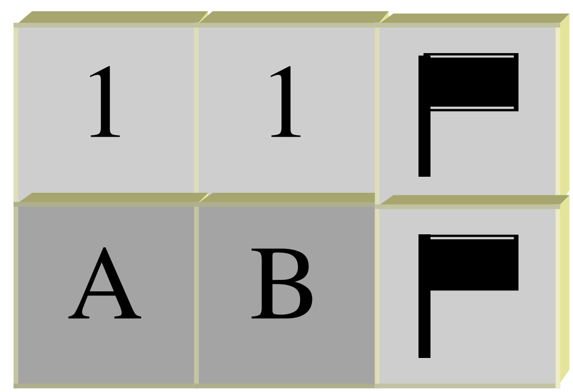
  $\quad + P(M(2, 1))$

  $\quad - 0$

  $P(M(1, 1)) = P(M(2, 1))$

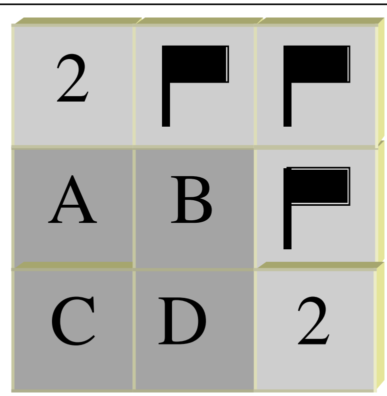
  $= .5$

Mine Sweeper



- **Exactly one of A and B must be a mine**
- **Exactly one of B and D must be a mine**

MinesLeft = 2



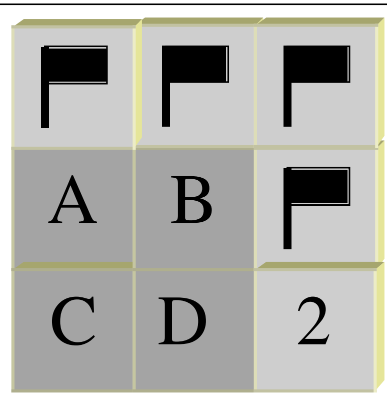
- **P(A) =**

$$\frac{\text{\# consistent patterns w/A}}{\text{\# consistent patterns}}$$

=1/2

- **Exactly one of B and D is a mine, so exactly one of A and C is a mine**

MinesLeft = 2



- **P(A) =**

 # consistent patterns w/A

 # consistent patterns

= 2/4 = 1/2

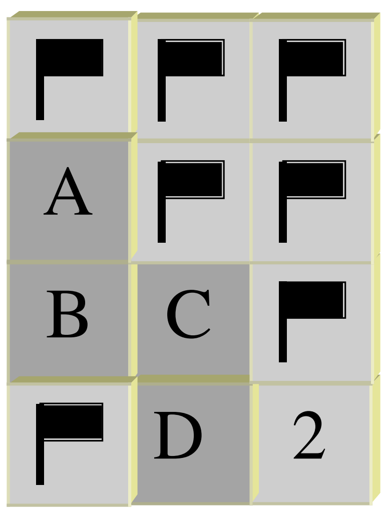
Joint Probability

- The state of a single situation / domain may be represented by a *set* of random variables.
- Joint probability distribution (or just "the joint"): probability distribution over all possible sets of values for the state variables.

Example of Joint Probability Distribution

Mine Sweeper

MinesLeft = 1



A	B	C	D	Probability
F	F	F	F	0
F	F	F	T	1/2
F	F	T	F	1/2
F	F	T	T	0
F	T	F	F	0
F	T	F	T	0
F	T	T	F	0
F	T	T	T	0
T	F	F	F	0
T	F	F	T	0
T	F	T	F	0
T	F	T	T	0
T	T	F	F	0
T	T	F	T	0
T	T	T	F	0
T	T	T	T	0

Diagnosis Problems: Inferring Causes from Effects

- **Reasoning from a single piece of evidence:**
  - We observe smoke (effect).
  - We want to know whether there is a fire (cause).
- **Reasoning from a multiple pieces of evidence:**
  - We observe that a patient has a fever (effect).
  - We observe that a patient has muscle aches (effect).
  - We want to know whether he has the flu (cause).

Available Knowledge

- **Probability of each effect given cause:**
  - $P(E|C)$.
- **Probability of each effect given absence cause:**
  - $P(E| \sim C)$
- **Probability of cause:**
  - $P(C)$
- **We don't assume knowledge of probability of cause given effect: $P(C|E)$.**

Using Bayes Rule for Diagnosis

- **Instantiating Bayes Rule:**

  $$P(C|E) = P(E|C)P(C) / P(E)$$

- **Normalization removes dependence on P(E):**

  $$P(C|E) = \frac{P(E|C)\,P(C)}{P(E|C)P(C) + P(E \mid \sim C)P(\sim C)}$$

Combining Pieces of Evidence with Bayesian Updating

- **Instantiating Bayes Rule:**

  $$P(C \mid E_1) = P(C)\, P(E_1 \mid C) / P(E_1)$$

- **Instantiating Bayes Rule w/ more Conditioning:**

  $$P(C \mid E_1 \wedge E_2) =$$

  $$P(C)\, P(E_1 \wedge E_2 \mid C) / P(E_1 \wedge E_2)$$

- **Applying Bayes Rule Again:**

  $$P(C \mid E_1 \wedge E_2) = \frac{P(C)\, P(E_1 \mid C)\, P(E_2 \mid C \wedge E_1)}{P(E_1)\;\; P(E_2 \mid E_1)}$$

Conditional Independence

- **Assumptions:**

  $$P(E\_1|C \wedge E\_2) = P(E\_1|C)$$

  $$P(E\_2|C \wedge E\_1) = P(E\_2|C)$$

  E_1 and E_2 are conditionally independent, given C

- **Therefore:**

  $$P(C|E\_1 \wedge E\_2) = \frac{P(C)\, P(E\_1|C)\, P(E\_2|C)}{P(E\_1)\, P(E\_2|E\_1)}$$

  $$P(C|E\_1 \wedge E\_2) = \text{alpha} * P(C)P(E\_1|C)P(E\_2|C)$$

  – **Where alpha is chosen by normalization.**

Joint Probability Distribution

- Choose a set of random variables: X_1,...,X_n.
- An atomic event (X_1=x_1 ,...,X_n=x_n) is an assignment of
- exactly one value to each varable.
- The joint probability distribution function
- P (X_1=x_1 ,...,X_n=x_n) assigns a probability to each atomic event.
- Encode as a table:
  - Size 2^N for boolean random variables.
  - Size D^N for general random variables.

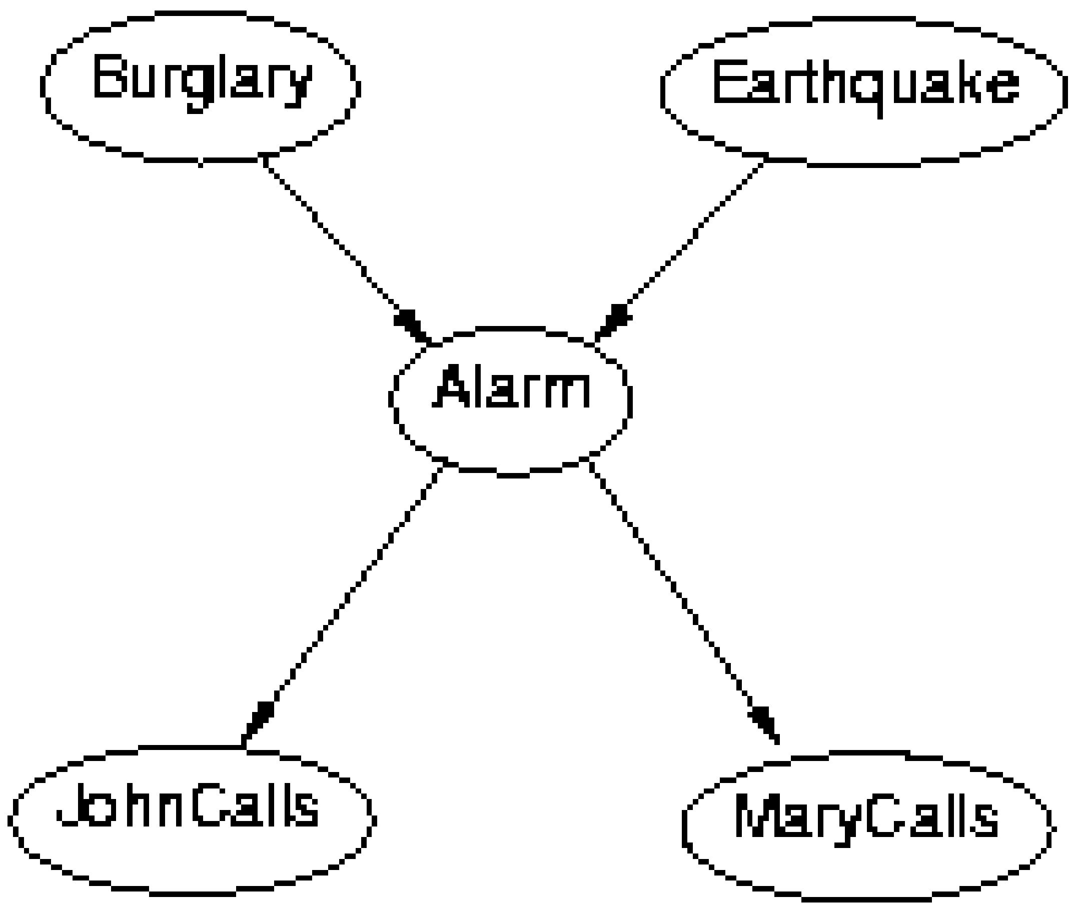
Using the Joint Probability Distribution

- **Difficulties:**
  - **Constructing the joint requires acquiring D^N numbers from an expert.**
  - **Computation with the joint requires lots of memory and CPU time.**
- **Opportunity:**
  - **The world is nearly decomposable.**
  - **Most things don't interact with each other.**
  - **Most joint entries are independent of each other.**

Belief Networks

- **Attempt to overcome difficulties of using the joint probability distribution**
- **A graphical representation of the joint probability distribution.**
- **A directed acyclic graph (DAG).**
- **Each node represents a random variable.**
- **Each arc represents a *direct* influence of one variable on another.**
- **Structure encodes conditional independence.**

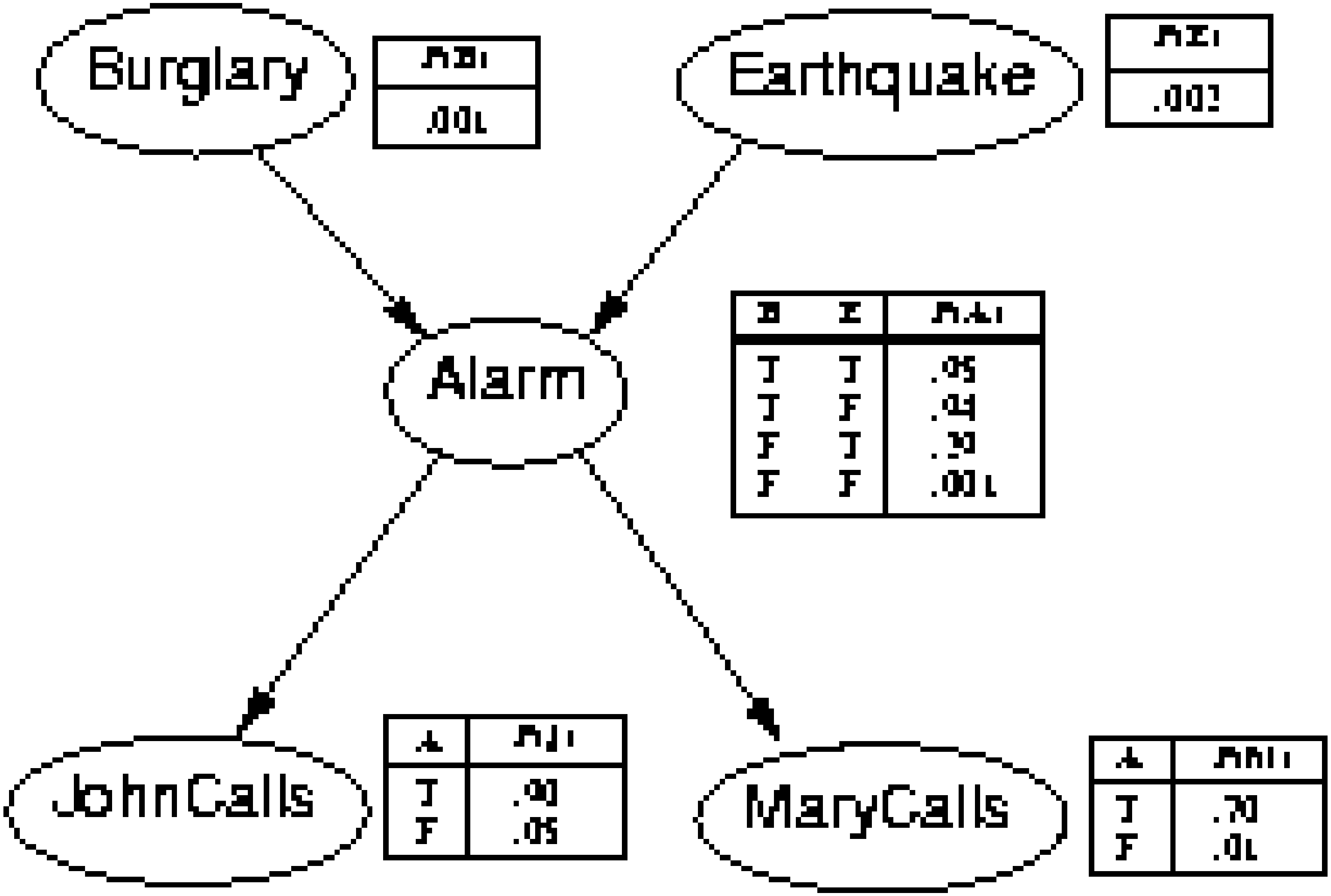
A Typical Belief Network



Semantics of Belief Networks

- A belief network encodes conditional probabilities.
- A conditional probability table is stored in each node.
- The table at the node for variable X_i stores the
- conditional probability: \newline
  - for each combination of values of V_i and values of V_j for all j \epsilon Parents(i).

Conditional Probabilities



Quantities in the Network

- $P(B) = 0.001$
- $P(E) = 0.002$

- $P(A \mid B \wedge E) = 0.95$
- $P(A \mid B \wedge \sim E) = 0.94$
- $P(A \mid \sim B \wedge E) = 0.29$
- $P(A \mid \sim B \wedge \sim E) = 0.001$

- $P(J \mid A) = 0.90$
- $P(J \mid \sim A) = 0.05$

- $P(M \mid A) = 0.70$
- $P(M \mid \sim A) = 0.01$

Extracting the Joint from a Belief Net

- **Boolean Case**
- **Let A be an atomic event.**
  - **Suppose:**

    $A == (X\_1 \wedge ... \wedge X\_n)$
  - **It follows that:**

    $P(A) = \prod_{I=1..n} P(X\_i | \wedge_{j \text{ in Parents}(i)} X\_j)$

Network Topology Constrains the Joint

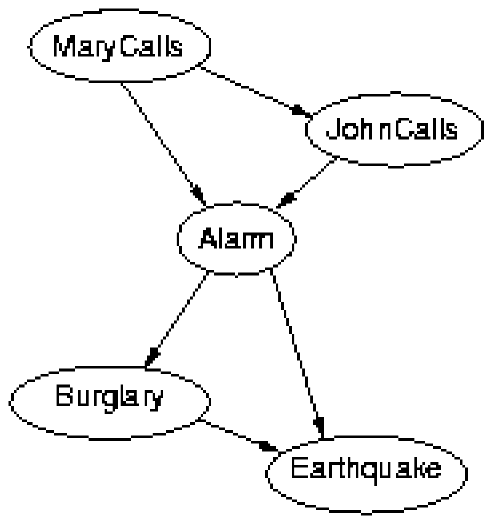
- **Conditional Independence Property:**

  Each variable is conditionally independent of its predecessors given its parents.

Constructing a Belief Network

- **1. Choose a set of variables.**
- **2. Choose an ordering of the variables.**
- **3 While (Some variables are left)**
- **a. Choose a variable, ``Xi''.**
- **b. Add a node for Xi to the network.**
- **c. Set Parents(i) to be the minimal set**
- **of nodes already in the network**
- **satisfying the conditional**
- **independence property.**
- **d. Define the conditional probability**
- **table for node Xi.**

Network Structure Depends on Order of Introduction



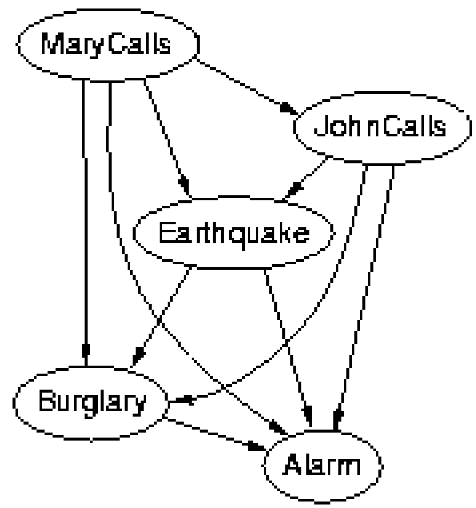
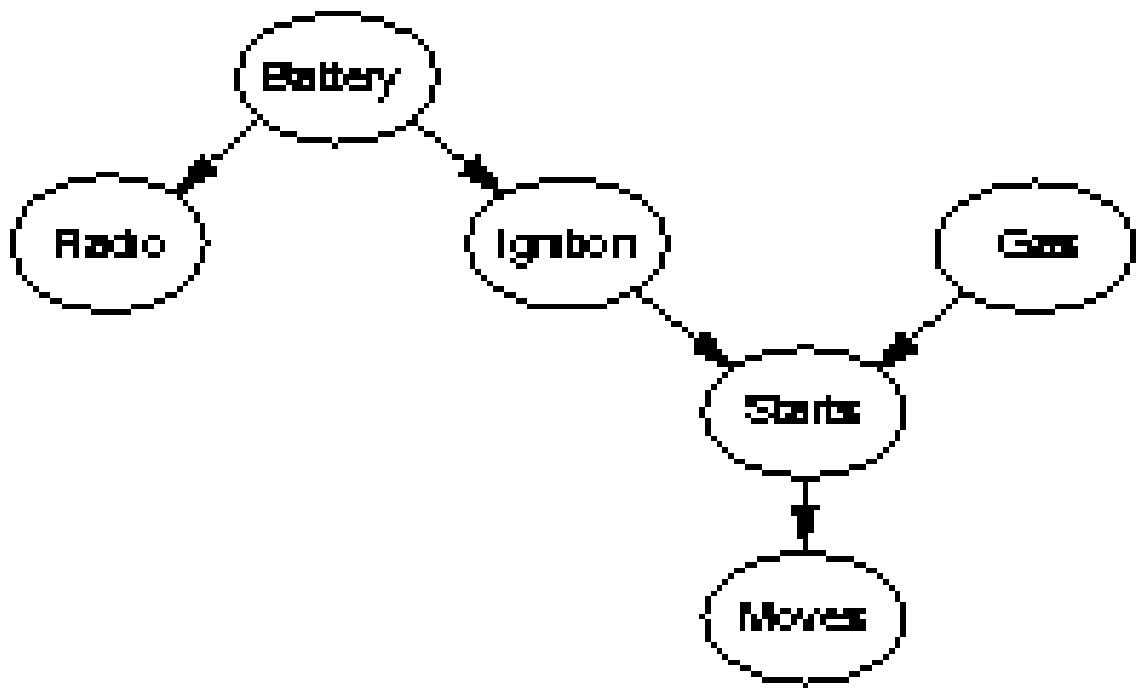


Illustration of Conditional Independence

– RADIO and GAS independent given IGNITION.
– RADIO and GAS independent given BATTERY.
– RADIO and GAS independent given nothing.



Patterns of Reasoning Handled by Belief Networks

- **Diagnostic Inference:**
  - **Given JohnCalls infer P(Burglary|JohnCalls).**
- **Causal Inference:**
  - **Given Burglary infer P(JohnCalls|Burglary).**
- **Intercausal Inference:**
  - **Given Alarm and Earthquake infer P(Burglary|Alarm ^ Earthquake)**
- **Mixed Inference.**

Terminology

- **Let `` B '' be ``Burglary.**
- **Let `` ~ E '' be ``Earthquake''.**
- **Let `` ~ A '' be ``Alarm''.**
- **Let `` ~ J '' be ``John Calls''.**
- **Let `` ~ M '' be ``Mary Calls''.**

Sample Calculations with the Belief Network

- Probability that there was a burglary given that the alarm goes off:
- P(B|A) = P(A|B) \frac P(B) P(A)
- Probability the alarm goes off given that there is a burglary:

P(A|B) = P(A ^ E|B) + P(A ^ ~E|B)

= P(A | E ^ B) P(E | B) + P(A |~E ^ B)P(~E | B)

= P(A |E ^ B) P(E) + P(A |~E ^ B) P(~E)

Sample Calculations with the Belief Network

Probability that the alarm goes off:

$$\begin{aligned}
P(A) = {} & P(A \wedge B \wedge E) \\
& + P(A \wedge B \wedge \sim E) \\
& + P(A \wedge \sim B \wedge E) \\
& + P(A \wedge \sim B \wedge \sim E)
\end{aligned}$$

$$\begin{aligned}
= {} & P(A \mid B \wedge E)\, P(B \wedge E) \\
& + P(A \mid B \wedge \sim E)\, P(B \wedge \sim E) \\
& + P(A \mid \sim B \wedge E)\, P(\sim B \wedge E) \\
& + P(A \mid \sim B \wedge \sim E)\, P(\sim B \wedge \sim E)
\end{aligned}$$

$$\begin{aligned}
= {} & P(A \mid B \wedge E)\, P(B)\, P(E) \\
& + P(A \mid B \wedge \sim E)\, P(B)\, P(\sim E) \\
& + P(A \mid \sim B \wedge E)\, P(\sim B)\, P(E) \\
& + P(A \mid \sim B \wedge \sim E)\, P(\sim B)\, P(\sim E)
\end{aligned}$$

Sample Calculations with the Belief Network

- **Probability that there is a burglary and Mary Calls:**
  - $P(B \wedge M) = P(B)\, P(M|B)$
  - $= P(B)\, P(M \wedge A \mid B)$
  - $+ P(B)\, P(M \wedge \sim A \mid B)$
  - $= P(B)\, P(M|A \wedge B)\, P(A|B)$
  - $+ P(B)\, P(M|\sim A \wedge B)\, P(\sim A|B)$
  - $= P(B)\, P(M|A)\, P(A|B)$
  - $+ P(B)\, P(M|\sim A)\, P(\sim A|B)$