Title: Probabilistic Reasoning Systems
AIMA: Chapter 15 (Section 15.1-15.3)

Introduction to Artificial Intelligence
CSCE 476-876, Spring 2000
`www.cse.unl.edu/~cse476`

Berthe Y. Choueiry
`choueiry@cse.unl.edu, (402)472-5444`

Outline

- Belief network: representation
- BN represents the joint probability distribution
  - Complete
  - Compact
- How to build/fill BN
- d-separation
- Nature of probabilistic reasoning

Belief networks a.k.a. $\left\{\begin{array}{l}\text{bayesian network}\\ \text{probabilistic network}\\ \text{causal networks}\\ \text{knowledge map}\end{array}\right.$

Representation:

- Random variables $\rightarrow$ nodes
- Directed links between some pairs of nodes
  Meaning of arrow $X \rightarrow Y$: "X has a direct influence on Y"
- One **conditional probability table** (CPT) per node
  quantifies effects of parents on node
- Network is a DAG (*i.e.*, no cycles)

Specified by: $\left\{\begin{array}{l}\text{Topology}\\ \text{CPT}\end{array}\right.$

Example: belief network

Burglar alarm at home $\left\{ \begin{array}{l} \text{responds to burglary} \\ \text{occasionally to minor earthquakes} \end{array} \right.$

Neighbors call $\left\{ \begin{array}{l} \text{Mary: loud music, may miss alarm} \\ \text{John: when hears alarm, but also sometimes phone} \end{array} \right.$

Assumption: Mary and John

don't observe burglary directly

don't feel minor earthquakes

Draw the belief network:

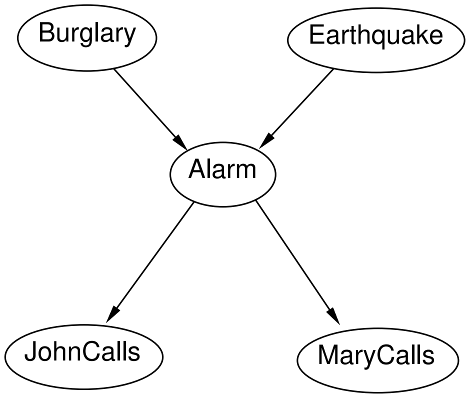
Random variables: Burglary, Earthquake, Alarm, JCalls, MCalls

Burglar alarm

Topology: $\begin{cases} \text{abstract knowledge base} \\ \text{represents structure of causal processes} \end{cases}$

Burglary, earthquakes directly affect alarm

John, Mary call depend only on alarm



Links: Alarm→ JCalls, Alarm→ MCalls summarize:

— Music, phone

— Potentially infinite other factors (dead battery, vacation, *etc.*)

Links model <u>laziness, ignorance</u>

Conditional probability table (1 CPT/node in BN)

Conditioning case: (a mini atomic event)

a possible combination of values for the parent nodes

Given a node n in BN, each row in CPT for n contains the conditional probability for each value of n for one conditioning case

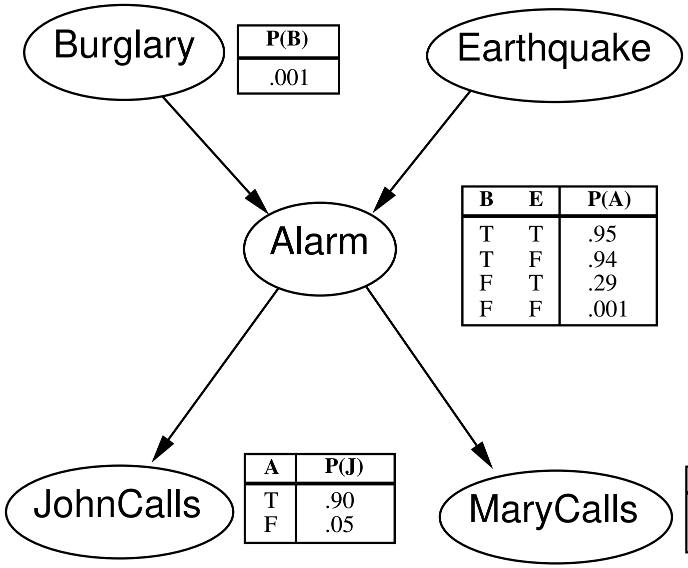
Burglary	Earthquake	$\mathbf{P}(Alarm \mid Burglary, Earthquake)$	
T	T	0.95	0.05
T	F	0.95	0.050
F	T	0.290	0.710
F	F	0.001	0.999

- n Boolean parents, 2^n entries in a CPT
- Exhaustive set of cases:

  each row in CPT sums to one 1 col is sufficient
- No parents? 1 row, prior probability

Belief net: example

Note: Topology, 1 CPT per node, rows per CPT



P(B)
.001

P(E)
.002

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

A	P(J)
T	.90
F	.05

A	P(M)
T	.70
F	.01

Semantics of BN

Two ways to understand a BN:

1. BN as a representation of the joint probability distribution (JPD)
2. BN as a encoding of a collection of independence statements

BN as joint prob. distribution (I)

Each entry in JPD: is a probability of an atomic event,

probability of a conjunction of assignments to each variable

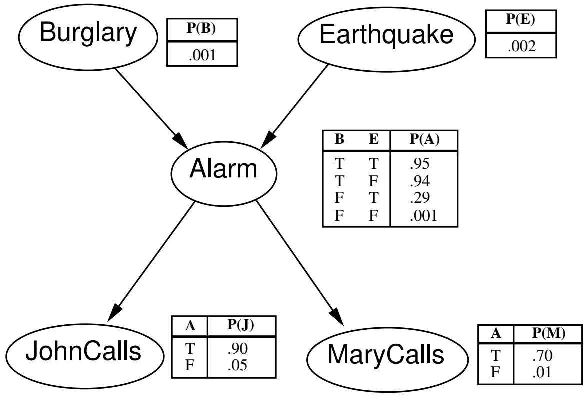
$$P(X_1 = x_1 \wedge \ldots \wedge X_n = x_n) = P(x_1, \ldots, x_n)$$

Its value in BN: $P(x_1, \ldots, x_n) = \prod_{i=1}^{n} P(x_i | Parents(X_i))$

It is thus represented by the product of appropriate elements of CPTs in BN

BN as joint prob. distribution (II)

Chain rule: $P(x_1, \ldots, x_n) = \prod_{i=1}^{n} P(x_i | Parents(X_i))$



Event: no burglary, no earthquake, alarm off, John & Mary call

$P(J \wedge M \wedge A \wedge \neg B \wedge \neg E)$

$= P(J|A)P(M|A)P(A|\neg B \wedge \neg E)P(\neg B)P(\neg E)$

$= .9 \times .7 \times .0001 \times .999 \times .998$

$= .00062$

BN as joint prob. distribution (III)

We know JPD can be used to answer **any** query about domain

We just saw BN is a representation of the JDP

Thus, BN can be used to answer **any** query about domain

How?

Use BN to compute all entries of JDP...

but, we can do better..

Constructing BN (I)

Using the product rule: $P(A \wedge B) = P(B|A)P(A)$

$$
\begin{aligned}
P(x_1, \ldots, x_n) &= P(x_n|x_{n-1}, \ldots, x_1)P(x_{n-1}, \ldots, x_1) \\
P(x_1, \ldots, x_n) &= P(x_n|x_{n-1}, \ldots, x_1)P(x_{n-1}|x_{n-2}, \ldots, x_1)P(x_{n-2}, \ldots, x_1) \\
&\vdots \\
P(x_1, \ldots, x_n) &= P(x_n|x_{n-1}, \ldots, x_1)P(x_{n-1}|x_{n-2}, \ldots, x_1) \ldots P(x_2|x_1)P(x_1) \\
&= \prod_{i=1}^{n} P(x_i|x_{i-1}, \ldots, x_1)
\end{aligned}
$$

Compare with: $P(x_1, \ldots, x_n) = \prod_{i=1}^{n} P(x_i|Parents(X_i))$ (1)

$\rightarrow$ Specification of the joint is equivalent to (1),
provided $Parents(X_i) \subseteq \{x_{i-1}, \ldots, x_1\}$

Constructing BN (II)

Product rule: $P(x_1, \ldots, x_n) = \prod_{i=1}^{n} P(x_i | x_{i-1}, \ldots, x_1)$
BN as JDP: $P(x_1, \ldots, x_n) = \prod_{i=1}^{n} P(x_i | Parents(X_i))$

$\rightarrow$ Build BN to satisfy this condition
label nodes in BN in any order consistent with
partial order implicit in the graph structure

$\rightarrow$ BN is a correct representation of a domain only if each node is
conditionally independent of its predecessors
in the node ordering, given its parents

Constructing BN (III)

Intuitively, parents of X_i should contain all nodes in $X_1, \ldots, X_{i-1}$ *that directly influence* X_i

Burglary example:

— MCalls is directly influenced by Alarm
not by Burglary or Earthquake

— Given Alarm, JCalls does not influence MCalls

Conditional independence statement:

$\mathbf{P}$(MCalls| JCalls, Alarm, Earthquake, Burglary) = $\mathbf{P}$ (.....)

Constructing BN (IV)

Procedure for incremental network construction:

1. Choose a set of variables X_i to describe domain
2. Choose an ordering for variables (very important)
3. While there are variables left
   (a) Pick a variable X_i and add a node in the BN for it
   (b) Set $Parents(X_i)$ to a minimal set of nodes already in the net <u>such that the conditional independence property</u> holds
   (c) Define the CPTable for X_i

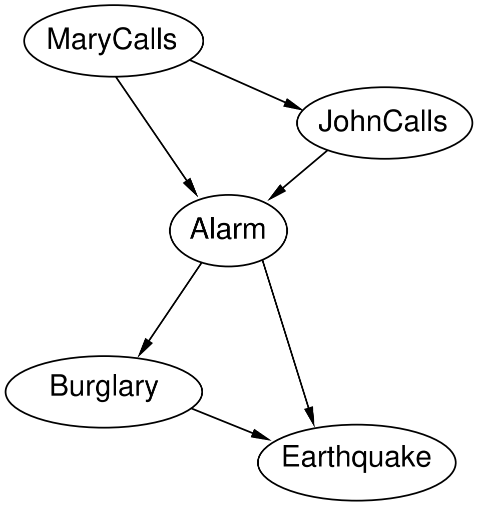
Note:

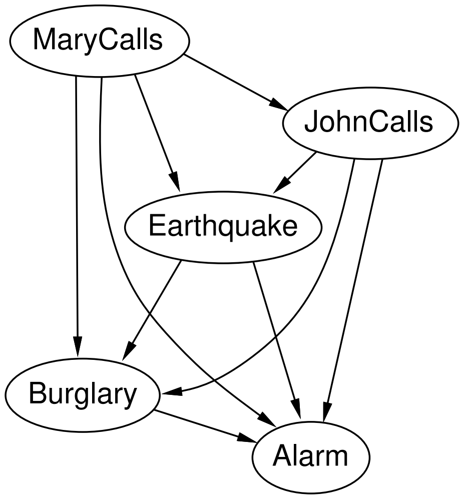
- Each node is connected only to previous nodes → BN is a DAG
- BN contains <u>**no redundant probability values**</u> (except 1 value in each row in each CPT)
  → Impossible for a Knowledge Engineer/domain expert to create a BN that violates the axioms of probability

Variable ordering (I)

Variables could be ordered according to:

- Causal rules: links from causes to symptoms
  Burglary$\rightarrow$ Alarm, Alarm$\rightarrow$ MCalls
- Diagnostic rules: links from symptoms to causes
  Alarm $\rightarrow$ Burglary, MCalls $\rightarrow$ Alarm
- Any other order





Variable ordering (II)

Different orders yield different DAGs

— more links, more rows in CPTs

— some unnatural P(E | Burglary $\wedge$ Alarm); require specifying additional dependencies between otherwise independent causes

— Worst case: same number as entries in JPD

Note: Any ordering represents the same joint distribution

> "If we stick to a <u>causal model</u> we end up having to specify fewer numbers, and the numbers will often be easier to come up with" (domain experts find it more natural)

Representation: JPD vs. BN

$$\text{BN provides a} \left\{ \begin{array}{l} \text{complete} \\ \text{non-redundant} \\ \text{compact} \end{array} \right\} \text{representation of a domain}$$

BN is compact, JDP has 2^n entries

Compactness: property of **locally structured, sparse domains**

- BN is sparse:
  Each node has at most k parent $\rightarrow$ CPT has 2^k rows
  BN has n nodes $\rightarrow$ n2^k entries to specify
  Example: 20 nodes, 5 parent per node, 640 numbers
  Full joint: over a million
- If BN fully connected:
  Each variable directly influenced by all others
  BN has as many numbers as full joint

Sparse networks

BN good at capturing structure of domain

- Domain has lots of structure: BN captures it
- Domain does not have lots of structure:
  Sometimes, there are some dependencies that strictly speaking should be included by adding a link, but are neglected because they are not worth the additional complexity

  Trade complexity for accuracy

  Example: If Earthquake, ¬JCalls and ¬MCalls even if Alarm

Filling CPTs (conditioning cases)

Difficult task even if BN is sparse $\begin{cases} \text{time} \\ \text{experience} \end{cases}$

Usually, $\exists$ canonical distributions: pick and plug

Deterministic nodes

$\rightarrow$ value determined by that of parents

e.g., logical (disjunction), numerical (max, difference)

'Noisy' logical relations

e.g. Noisy-OR, generalization of logical OR-relation:

FOL: $Fever \Leftrightarrow Cold \lor Flu \lor Malaria$

Noisy-OR logical relations

FOL: $Fever \Leftrightarrow Cold \vee Flu \vee Malaria$

Noisy-OR adds uncertainty under 3 assumptions:

1. Each cause has an independent chance of causing effect
2. All possible causes are listed
   (one could add a leak node to cover other all other causes)
3. Independence of negation of causes to effect
   Whatever inhibits Cold from causing Fever is independent from whatever inhibits Flu from causing fever

   Inhibitors are not nodes, but noise parameters:
   $P(Fever|Cold) = 0.4, P(Fever|Flu) = 0.8, P(Fever|Malaria) = 0.9$
   then noise parameters are: 0.6, 0.2, 0.1

Filling CPT of Noisy-OR

$P(Fever|Cold) = .4, P(Fever|Flu) = .8, P(Fever|Malaria) = .9$
Noise parameters (NP) are: .6, .2, .1

— If no Parent is true, child is False 100% certainty

— If exactly 1 parent is True, child is False with NP of this parent

— Otherwise, probability child is False is product of NP of parents that are True

Cold	Flu	Malaria	P(Fever)	P(¬Fever)
F	F	F		
F	F	T		
F	T	F		
F	T	T		
T	F	F		
T	F	T		
T	T	F		
T	T	T		

Semantics of BN

Two ways to understand a BN:

1. BN as a representation of the joint probability distribution (JPD) $\surd$

2. BN as a encoding of a collection of independence statements $\longleftarrow$

Conditional independence relations in BN

So far:

- we argued that a BN expresses the conditional independence of a node and its predecessors, given its parent
- we used this independence to design a construction method for networks

How about more general conditional independence?

Given in a BN, $\begin{cases} \text{a set of nodes } X \\ \text{a set of nodes } Y \\ \text{a set of evidence nodes } E \end{cases}$

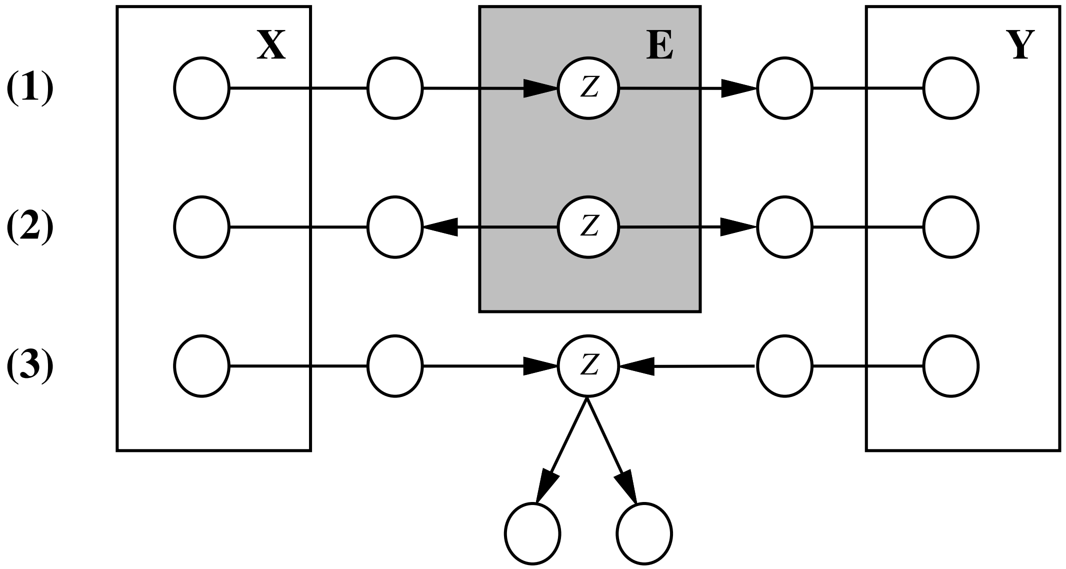
Question: determine if X and Y are conditionally independent given E?

Answer: <u>direction-dependent separation</u> or <u>d-separation</u>

Note: useful for designing inference algorithm

D-separation (I)

"If every undirected path from a node in X to a node in Y is d-separated by E, then X and Y are conditionally independent given E"



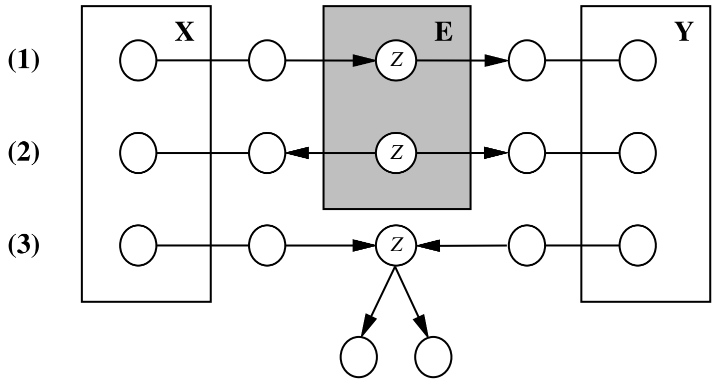
D-separation (II)

Paths between nodes in X to nodes in Y are blocked given E

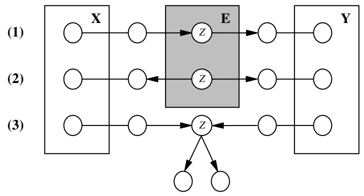
Path is blocked if $\exists$ node Z on the path such that:

1. $Z \in E$, and Z has 1 arrow in and 1 arrow out
2. $Z \in E$, and Z has both arrows leading out
3. Neither Z nor any descendant of Z is in E
   AND both arrows lead in to Z

Proof is omitted



Example D-separation

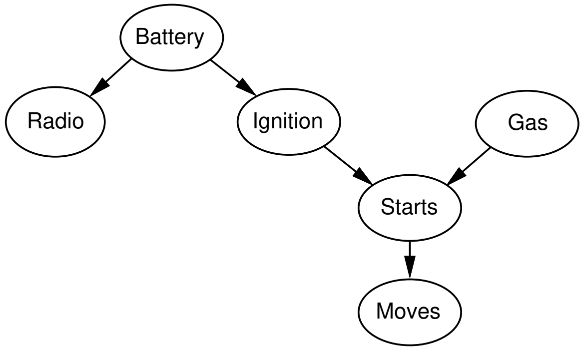


Car, X and Y are independent

1. $X = \{\text{Gas}\}$, $Y = \{\text{Radio}\}$, $E = \{\text{SparkPlugs fires}\}$
2. $X = \{\text{Gas}\}$, $Y = \{\text{Radio}\}$, $E = \{\text{Battery works}\}$
3. $X = \{\text{Gas}\}$, $Y = \{\text{Radio}\}$, $E = \{\}$

Dependent: $X = \{\text{Gas}\}$, $Y = \{\text{Radio}\}$, $E = \{\text{Car Starts}\}$

Dependent: $X = \{\text{Gas}\}$, $Y = \{\text{Radio}\}$, $E = \{\text{Moves}\}$



Inference in BN

Task: $\mathbf{P}$(Query | Evidence)

Given exact values for some evidence variables, compute posterior probability distribution for a set of query variables

e.g., $\mathbf{P}$ (Burglary | JCalls $\wedge$ MCalls)?

Agent:

— gets values for evidence from its percepts (or other reasoning),

— asks about the possible values of other variables

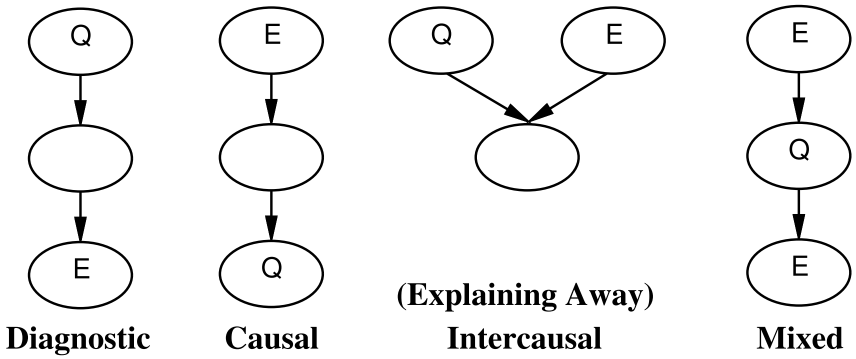
— so that it can decide what actions to take

Agent needs 2 functions:

$$\begin{cases} \text{BN-Tell: to add evidence} \\ \text{BN-Ask: to compute posterior probability distribution for Query} \end{cases}$$

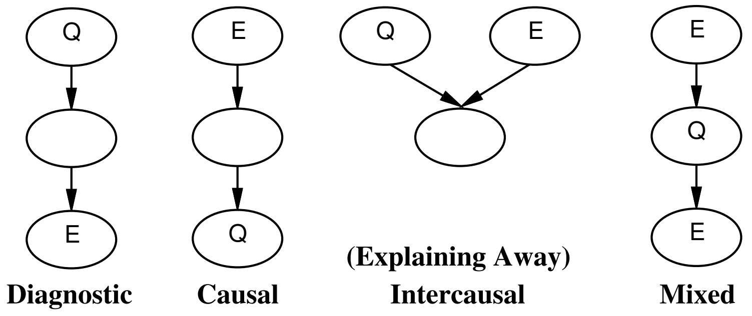
Nature of probabilistic inferences

- Diagnostic inference
- Causal inferences
- Intercausal inferences
- Mixed inferences: combining 2 or more of above



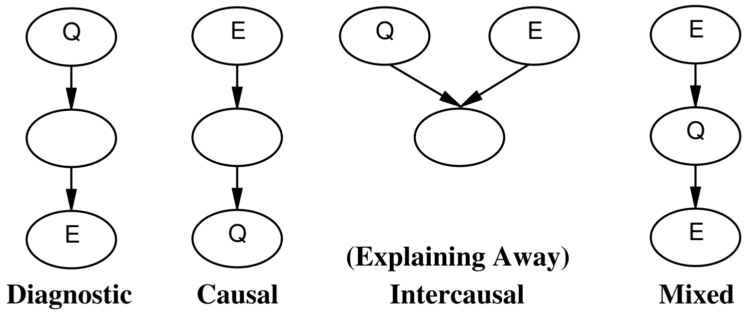
Diagnostic inference: from effects to causes

Given JCalls, infer P(Burglary | JCalls)



Causal inferences: from causes to effects

Given Burglary, P(JCalls | Burglary)? P(MCalls | Burglary)?

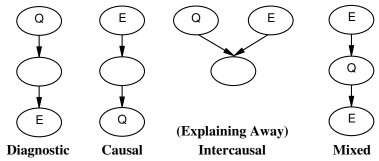


Intercausal inferences: between causes of a common effect

Given Alarm, P(Burglary | Alarm)= 0.376
Add evidence Earthquake,

P(Burglary | Alarm $\wedge$ Earthquake)= 0.003

Even if Burglaries & Earthquakes are independent!!
Presence of one makes the other less likely: explaining away



New cause explains away a symptom, probability of first cause goes down. Typical for probabilisitc reasoning. Impossible to generate with logical reasoning.

Mixed inferences: combining 2 or more of above

Diagnostic+Causal inference: P(Alarm | JCalls$\wedge\neg$Earthquake)

Intercausal+Diagnostic: P(Burglary | JCalls$\wedge\neg$Earthquake)