Reasoning with Joint Probability

Example: Alarm

- Story: In LA, burglary and earthquake are not uncommon. They both can cause alarm. In case of alarm, two neighbors John and Mary may call.
- Variables: Burglary (B), Earthquake (E), Alarm (A), JohnCalls (J), MaryCalls (M).
- Problem: Estimate the probability of a burglary based on who has or has not called.
- Problem can be solved if *joint probability* is available.

$$P(B, E, A, J, M)$$

Reasoning with Joint Probability

B	E	A	J	M	Prob	B	E	A	J	M	Prob
y	y	y	y	y	.00001	n	y	y	y	y	.0002
y	y	y	y	n	.000025	n	y	y	y	n	.0004
y	y	y	n	y	.000025	n	y	y	n	y	.0004
y	y	y	n	n	.00000	n	y	y	n	n	.0002
y	y	n	y	y	.00001	n	y	n	y	y	.0002
y	y	n	y	n	.000015	n	y	n	y	n	.0002
y	y	n	n	y	.000015	n	y	n	n	y	.0002
y	y	n	n	n	.00000	n	y	n	n	n	.0002
y	n	y	y	y	.00001	n	n	y	y	y	.0001
y	n	y	y	n	.000025	n	n	y	y	n	.0002
y	n	y	n	y	.000025	n	n	y	n	y	.0002
y	n	y	n	n	.00000	n	n	y	n	n	.0001
y	n	n	y	y	.00001	n	n	n	y	y	.0001
y	n	n	y	n	.00001	n	n	n	y	n	.0001
y	n	n	n	y	.00001	n	n	n	n	y	.0001
y	n	n	n	n	.00001	n	n	n	n	n	.9969

Reasoning with Joint Probability

- $P(B{=}y|M{=}y)$?
- Compute *marginal probability*:

$$P(B, M) = \sum_{E,A,J} P(B, E, A, J, M)$$

B	M	Prob
y	y	.000115
y	n	.000085
n	y	.0015
n	n	.9983

-

$$P(B{=}y|M{=}y) = \frac{P(B{=}y, M{=}y)}{P(M{=}y)} = \frac{.000115}{.000115 + .0015} = 0.07$$

Reasoning with Joint Probability

Advantages:

- Clear semantics
- In theory, can perform arbitrary inference among the variables.

Difficulties: Complexity

- In the Alarm example:
  - 31 numbers needed,
  - quite unnatural to assess: e.g.

$$P(B = y, E = y, A = y, J = y, M = y)$$

  - Many additions in inference.
- In general, $P(X_1, X_2, \ldots, X_n)$ needs at least $2^n - 1$ numbers to specify the joint probabilities.
  - Knowledge acquisition (complex, unnatural)
  - Storage
  - Inference

Reasoning with Joint Probability

Solution:

- Use product rule:

$$\begin{aligned} P(B,E,A,J,M) &= P(B)P(E,A,J,M|B) \\ &= P(B)P(E|B)P(A,J,M|B,E) \\ &= \ldots \\ &= P(B)P(E|B)P(A|B,E) \\ &\quad P(J|B,E,A)P(M|B,E,A,J). \end{aligned}$$

- Conditional independences:
  - $P(E|B) = P(E)$.
  - $P(J|B,E,A) = P(J|A)$.
  - $P(M|B,E,A,J) = P(M|A)$.
- Thus

$$P(B,E,A,J,M) = P(B)P(E)P(A|B,E)P(J|A)P(M|A).$$

- Probabilities:

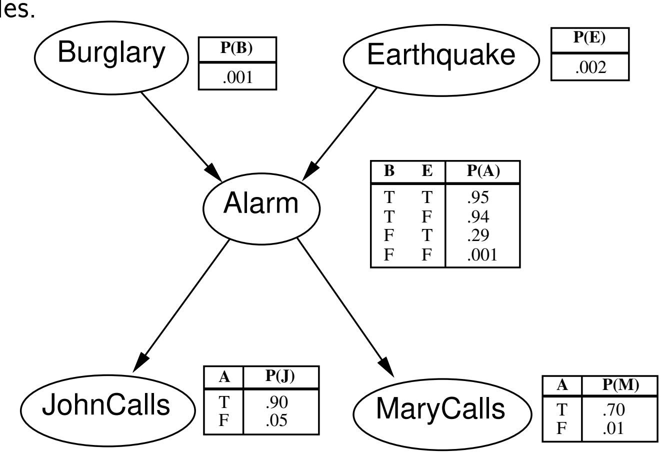
B	$P(B)$	E	$P(E)$	A	J	$P(J\|A)$	A	M	$P(M\|A)$
y	.001	y	.002	y	y	.9	y	y	.7
				n	y	.05	n	y	.01

B	E	A	$P(A\|B,E)$	B	E	A	$P(A\|B,E)$
y	y	y	.95	n	y	y	.29
y	n	y	.94	n	n	y	.001

- Only 10 numbers are needed. Those numbers are natural to assess.

Belief Networks

- Draw an arc to each variable from each of its conditioning variables.



- Attach to each variable its conditional probability table (CPT).
- Result: A *belief network*.
- Also know as Bayesian networks, Bayesian belief networks, probabilistic influence diagrams.

Belief Networks

- A *belief network* is a compact graphical representation of the joint probability distribution of the random variables.
- A belief network is a directed acyclic graph (DAG):
  - Nodes: variables
  - Links: direct probabilistic dependencies between variables
  - Node attributes: conditional probability tables (CPT)
- As a decomposable representation of the joint probability distribution:
  - The *global* joint probability is expressed as the product of *local* conditional probabilities:

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

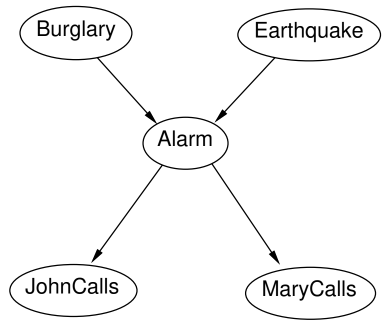
    where $Parents(X_i)$ is the set of parent nodes of X_i.
  - This view can help to design the network construction procedure.
- Example: Calculate the probability of the event that the alarm has sounded but neither a burglary nor and earthquake has occurred, and both John and Mary call.

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

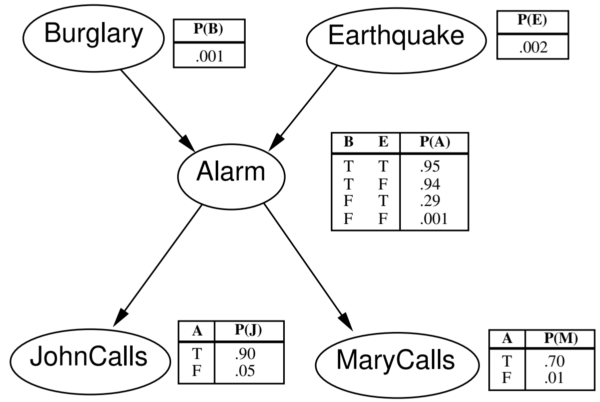
Belief Networks

Two levels:

- Qualitative level: Network structure



  - A depends on B and E.
  - J and M are conditionally independent given the event A.
- Numerical level: Conditional probabilities



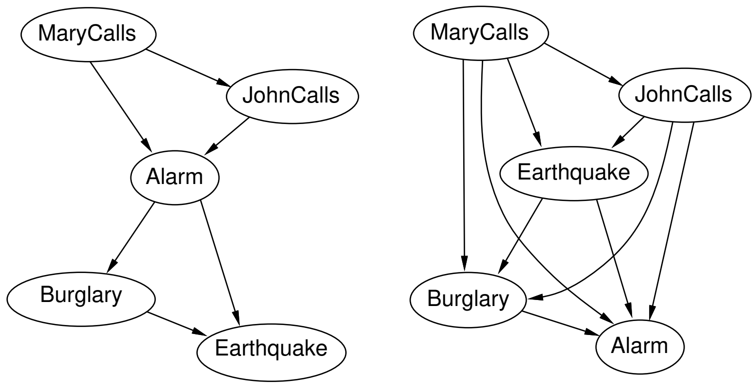
  - The CPT for a Boolean variable with n Boolean parents contains 2^n rows.
  - B and E have no parent nodes. The conditional probabilities thus degenerate to prior probabilities.

Network Construction Procedure

1. Choose the set of relevant variables X_i that describe the domain.
2. Choose an ordering for the variables.
3. While there are variables left:
   (a) Pick a variable X_i and add the corresponding node to the network under construction.
   (b) Set $Parents(X_i)$ to some minimal set of nodes already in the network such that the following conditional independence property holds:
   $$P(X_i|X_1,\ldots,X_{i-1}) = P(X_i|Parents(X_i))$$
   where $Parents(X_i) \subseteq \{X_1,\ldots,X_{i-1}\}$.
   (c) Define the CPT for X_i.

Node Ordering

- The resulting network structure depends on the order:
  - Example 1 (left): in the order M, J, A, B, E
  - Example 2 (right): in the order M, J, E, B, A



- Which order?

1. Naturalness of probability assessment.

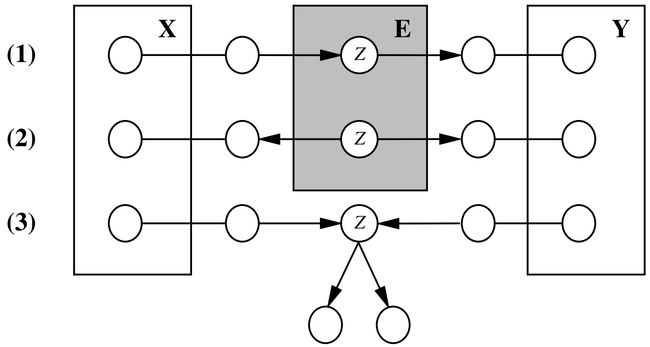
   The order M, J, E, B, A is not good because, for instance, $P(B|J,M,E)$ is unnatural and hence difficult to assess directly.
2. Minimize number of arcs.

   The order M, J, E, B, A is bad because there are too many arcs. The order B, E, A, J, M is good.
3. Use causal relationships: causes come before their effects.

   The order M, J, E, B, A is not good because, for instance, M and J are effects of A but come before A.

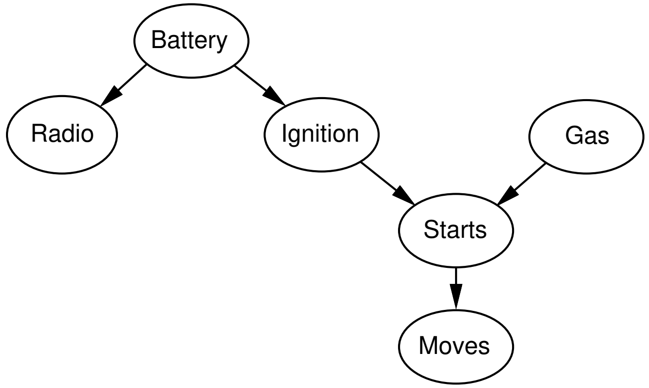
Conditional Independence in Belief Networks

- To determine whether a set of nodes X is independent of another set Y, given a set of evidence nodes E.
- 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.
- A set of nodes E d-separates two sets of nodes X and Y if every undirected path from a node in X to a node in Y is blocked given E.
- A path is blocked given a set of nodes E if there is a node Z on path for which one of three conditions holds:
  1. Z is in E and Z has one arrow on the path leading in and one arrow out.
  2. Z is in E and Z has both path arrows leading out.
  3. Neither Z nor any desendant of Z is in E, and both path arrows lead in to Z.
- Figure illustrating the three cases:



Conditional Independence: An Example

Example to illustrate the three cases of conditional independence:



1. Whether there is Gas in the car and whether the car Radio plays are independent given evidence about whether the SparkPlugs fire.
2. Gas and Radio are independent if given the Battery works.
3. Gas and Radio are independent given no evidence at all. But they are dependent given evidence about whether the car Starts.

Causality and Bayesian Networks

- Arcs in a Bayesian network can usually (but not always) be interpreted as indicating cause-effect relationships. They are from causes to effects (*causal network*).
- Making use of cause-effect relationships in Bayesian network construction:
  - Choose a set of variables that describes the domain.
  - Draw an arc to a variable from each of its DIRECT causes.
  - Assess the conditional probability of each node given its parents.

Examples

1. Poole's example:

   tampering, fire, smoke, alarm, people-leaving.

2. Africa visit

   Africa-visit, TB, smoke, lung-cancer, bronchitis, dyspnea, positive-X-yay.

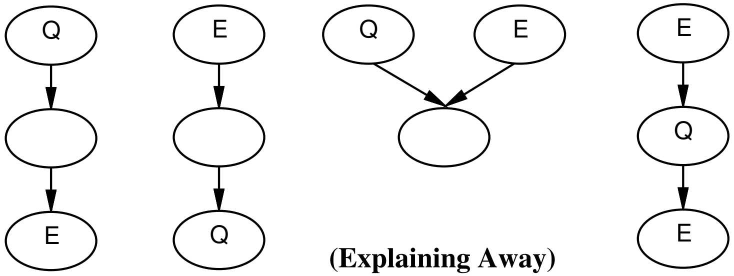
3. 1-bit adder

Inference

- Any node can serve as either a *query* variable or an *evidence* variable.
- Four inference types:
  - Diagnostic inference:
    - From effects to causes.
    - Example: $P(Burglary|JohnCalls)$
  - Causal inference:
    - From causes to effects.
    - Example: $P(JohnCalls|Burglary)$
  - Intercausal inference:
    - Between causes of a common effect.
    - Example: $P(Burglary|Alarm, Earthquake)$
      (Even though burglaries and earthquakes are independent, the presence of one makes the other less likely.)
  - Mixed inference:
    - Combining two or more of the above.
    - Example: $P(Alarm|JohnCalls, \neg Earthquake)$
      (Simultaneous use of diagnostic and causal inferences)
    - Example: $P(Burglary|JohnCalls, \neg Earthquake)$
      (Simultaneous use of diagnostic and intercausal inferences)

Inference (cont'd)

- Examples:



Diagnostic **Causal** **(Explaining Away) Intercausal** **Mixed**

- There exist efficient algorithms for all types of inference.