Bayesian Networks

Material used

– Halpern: Reasoning about Uncertainty. Chapter 4

– Stuart Russell and Peter Norvig: Artificial Intelligence: A Modern Approach

1 Random variables

2 Probabilistic independence

3 Belief networks

4 Global and local semantics

5 Constructing belief networks

6 Inference in belief networks

1 Random variables

- Suppose that a coin is tossed five times. What is the total number of heads?

- Intuitively, it is a *variable* because its value varies, and it is *random* because its value is unpredictable in a certain sense

- Formally, a random variable is neither random nor a variable

Definition 1: A random variable X on a sample space (set of possible worlds) W is a function from W to some range (e.g. the natural numbers)

Example

- A coin is tossed five times: $W = \{h,t\}^5$.
- $NH(w) = |\{i: w[i] = h\}|$ (number of heads in seq. w)
- $NH(hthht) = 3$
- Question: what is the probability of getting three heads in a sequence of five tosses?
- $\mu(NH = 3) =_{def} \mu(\{w: NH(w) = 3\})$
- $\mu(NH = 3) = 10 \cdot 2^{-5} = 10/32$

Why are random variables important?

- They provide a tool for structuring possible worlds
- A world can often be completely characterized by the values taken on by a number of random variables

- Example: $W = \{h,t\}^5$, each world can be characterized
  - by 5 random variables $X_1, \ldots X_5$ where X_i designates the outcome of the ith tosses: $X_i(w) = w[i]$
  - an alternative way is in terms of Boolean random variables, e.g. H_i: $H_i(w) = 1$ if $w[i] = h$, $H_i(w) = 0$ if $w[i] = t$.
  - use the random variables $H_i(w)$ for constructing a new random variable that expresses the number of tails in 5 tosses

2 Probabilistic Independence

- If two events U and V are independent (or unrelated) then learning U should not affect he probability of V and learning V should not affect the probability of U.

Definition 2: U and V are absolutely independent (with respect to a probability measure μ) if $\mu(V) \neq 0$ implies $\mu(U|V) = \mu(U)$ and $\mu(U) \neq 0$ implies $\mu(V|U) = \mu(V)$

Fact 1: the following are equivalent

a. $\mu(V) \neq 0$ implies $\mu(U|V) = \mu(U)$

b. $\mu(U) \neq 0$ implies $\mu(V|U) = \mu(V)$

c. $\mu(U \cap V) = \mu(U)\,\mu(V)$

Absolute independence for random variables

Definition 3: Two random variables X and Y are absolutely independent (with respect to a probability measure μ) iff for all $x \in$ Value(X) and all $y \in$ Value(Y) the event $X = x$ is absolutely independent of the event $Y = y$.
Notation: $I_\mu(X,Y)$

Definition 4: n random variables $X_1 \ldots X_n$ are absolutely independent iff for all i, $x_1, \ldots, x_n$, the events $X_i = x_i$ and $\cap_{j \neq i}(X_i = x_i)$ are absolutely independent.

Fact 2: If n random variables $X_1 \ldots X_n$ are absolutely independent then $\mu(X_1 = x_1, X_n = x_n) = \Pi_i \, \mu(X_i = x_i)$.

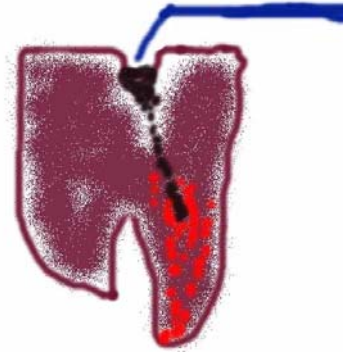Absolute independence is a very strong requirement, seldom met

Conditional independence: example

Example: *Dentist problem* with three events:
  Toothache (I have a toothache)
  Cavity (I have a cavity)
  Catch (steel probe catches in my tooth)

- If I have a cavity, the probability that the probe catches in it does not depend on whether I have a toothache
- i.e. *Catch* is conditionally independent of *Toothache* given *Cavity*: $I_\mu(Catch, Toothache|Cavity)$
- $\mu(Catch|Toothache \cap Cavity) = \mu(Catch|Cavity)$

Conditional independence for events

Definition 5: A and B are conditionally independent given C if

$\mu(B \cap C) \neq 0$ implies $\mu(A|B \cap C) = \mu(A|C)$ and
$\mu(A \cap C) \neq 0$ implies $\mu(B|A \cap C) = \mu(B|C)$

Fact 3: the following are equivalent if $\mu(C) \neq 0$

  a. $\mu(A|B \cap C) \neq 0$ implies $\mu(A|B \cap C) = \mu(A|C)$

  b. $\mu(B|A \cap C) \neq 0$ implies $\mu(B|A \cap C) = \mu(B|C)$

  c. $\mu(A \cap B|C) = \mu(A|C)\,\mu(B|C)$

Conditional independence for random variables

Definition 6: Two random variables X and Y are conditionally independ. given a random variable Z iff for all $x \in Value(X)$, $y \in Value(Y)$ and $z \in Value(z)$ the events $X = x$ and $Y = y$ are conditionally independent given the event $Z = z$.
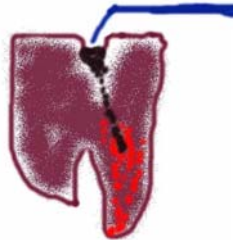*Notation:* $I_\mu(X,Y|Z)$

Important Notation: Instead of

(*) $\mu(X=x \cap Y=y | Z=z) = \mu(X=x|Z=z) \, \mu(Y=y|Z=z)$

we simply write

(**) $\mu(X,Y|Z) = \mu(X|Z) \, \mu(Y|Z)$

Question: How many equations are represented by (**)?

Dentist problem with random variables

- Assume three binary (Boolean) random variables *Toothache, Cavity*, and *Catch*

- Assume that *Catch* is conditionally independent of *Toothache* given *Cavity*

- The full joint distribution can now be written as
  $\mu(\textit{Toothache, Catch, Cavity}) =$
  $\mu(\textit{Toothache, Catch}|\textit{Cavity}) \cdot \mu(\textit{Cavity}) =$
  $\mu(\textit{Toothache}|\textit{Cavity}) \cdot \mu(\textit{Catch}|\textit{Cavity}) \cdot \mu(\textit{Cavity})$

- In order to express the full joint distribution we need 2+2+1 = 5 independent numbers instead of 7! 2 are removed by the statement of conditional independence:
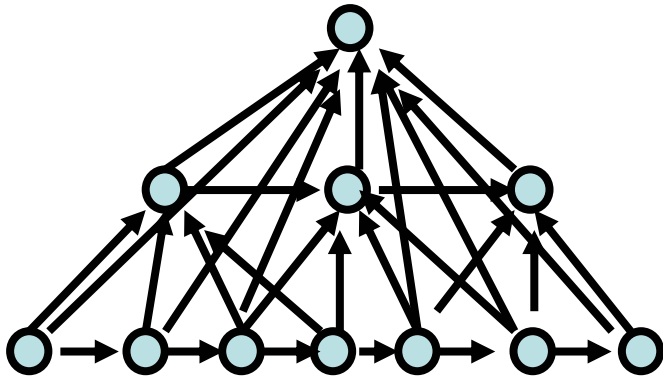
  $\mu(\textit{Toothache, Catch}|\textit{Cavity}) = \mu(\textit{Toothache}|\textit{Cavity}) \cdot \mu(\textit{Catch}|\textit{Cavity})$

3 Belief networks

- A simple, graphical notation for conditional independence assertions and hence for compact specification of full joint distribution.

- Syntax:
    - a set of nodes, one per random variable
    - a directed, acyclic graph (link ≈ "directly influences")
    - a conditional distribution for each node given its parents
        $\mu(X_i|\text{Parents}(X_i))$
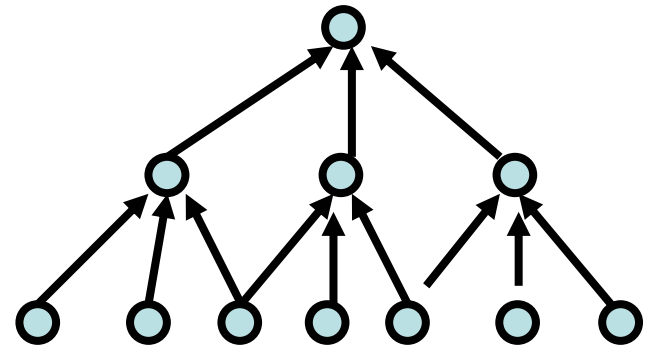- Conditional distributions are represented by conditional probability tables (CPT)

The importance of independency statements

n binary nodes,
fully connected

n binary nodes
each node max. 3 parents

$2^n -1$ independent numbers

less than $2^3 \cdot n$
independent numbers
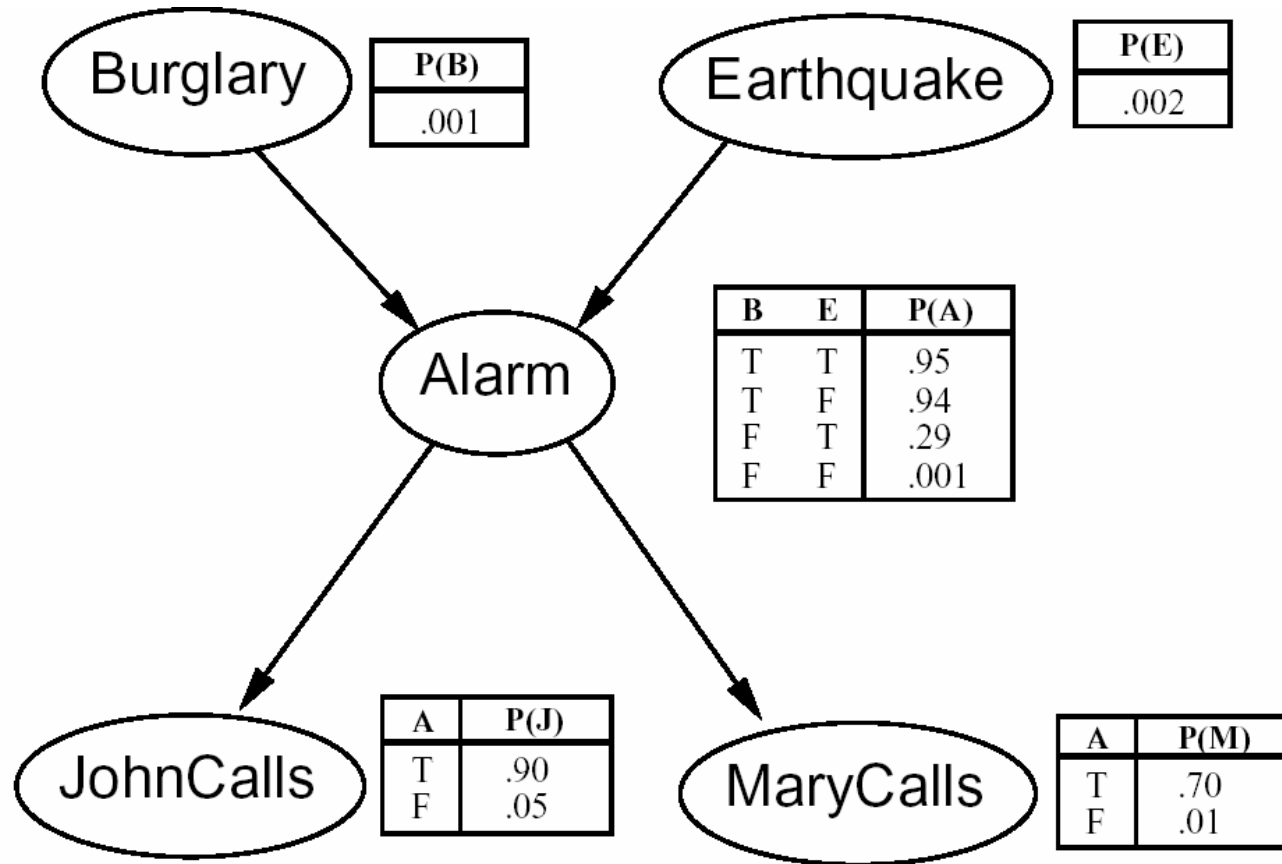
The earthquake example

- You have a new burglar alarm installed
- It is reliable about detecting burglary, but responds to minor earthquakes
- Two neighbors (John, Mary) promise to call you at work when they hear the alarm
  - John always calls when hears alarm, but confuses alarm with phone ringing (and calls then also)
  - Mary likes loud music and sometimes misses alarm!
- Given evidence about who has and hasn't called, estimate the probability of a burglary

The network

I´m at work, John calls to say my alarm is ringing, Mary doesn´t call. Is there a burglary?

5 Variables

network topol-ogy reflects *causal* knowledge

4 Global and local semantics

- *Global semantics* (corresponding to Halpern´s quantitative Bayesian network) defines the full joint distribution as the product of the local conditional distributions

- For defining this product, a linear ordering of the nodes of the network has to be given: $X_1 \dots X_n$

- $\mu(X_1 \dots X_n) = \prod_{i=1}^{n} \mu(X_i | \text{Parents}(X_i))$

- ordering in the example: B, E, A, J, M

- $\mu(J \cap M \cap A \cap \neg B \cap \neg E) =$

  $\mu(\neg B) \cdot \mu(\neg E) \cdot \mu(A | \neg B \cap \neg E) \cdot \mu(J|A) \cdot \mu(M|A)$

Local semantics

- *Local semantics* (corresponding to Halpern´s qualitative Bayesian network) defines a series of statements of conditional independence

- Each node is conditionally independent of its nondescendants given its parents: $I_\mu(X, \text{Nondescentents}(X)|\text{Parents}(X))$

- Examples

  - $X \rightarrow Y \rightarrow Z$ $I_\mu(X, Y)$? $I_\mu(X, Z)$?
  - $X \leftarrow Y \rightarrow Z$ $I_\mu(X, Z|Y)$?
  - $X \rightarrow Y \leftarrow Z$ $I_\mu(X, Y)$? $I_\mu(X, Z)$?

The chain rule

- $\mu(X, Y, Z) = \mu(X) \cdot \mu(Y, Z \mid X) = \mu(X) \cdot \mu(Y \mid X) \cdot \mu(Z \mid X, Y)$

- In general: $\mu(X_1, \ldots, X_n) = \prod_{i=1}^{n} \mu(X_i \mid X_1, \ldots, X_{i-1})$

- a linear ordering of the nodes of the network has to be given: $X_1, \ldots, X_n$

- The chain rule is used to prove

  the equivalence of local and global semantics

Local and global semantics are equivalent

- If a local semantics in form of the independeny statements is given, i.e.
  $I_\mu(X, \text{Nondescendants}(X)|\text{Parents}(X))$ for each node X of the network,

  then the global semantics results:
  $\mu(X_1 \ldots X_n) = \Pi^n_{i=1} \mu(X_i|\text{Parents}(X_i))$,
  and *vice versa.*

- For proving local semantics $\rightarrow$ global semantics, we assume an ordering of the variables that makes sure that parents appear earlier in the ordering: X_i parent of X_j then $X_i < X_j$

Local semantics → global semantics

- $\mu(X_1, \ldots, X_n) = \Pi^n_{i=1}\ \mu(X_i | X_1, \ldots, X_{i-1})$ chain rule

- $\text{Parents}(X_i) \subseteq \{X_1, \ldots, X_{i-1}\}$

- $\mu(X_i | X_1, \ldots, X_{i-1}) = \mu(X_i | \text{Parents}(X_i), \text{Rest})$

- local semantics: $I_\mu(X, \text{Nondescendants}(X) | \text{Parents}(X))$

- The elements of Rest are nondescendants of X_i, hence we can skip Rest

- Hence, $\mu(X_1 \ldots X_n) = \Pi^n_{i=1}\ \mu(X_i | \text{Parents}(X_i))$,

5 Constructing belief networks

Need a method such that a series of locally testable assertions of conditional independence guarantees the required global semantics

1. Chose an ordering of variables $X_1, \ldots, X_n$
2. For i = 1 to n
   add X_i to the network
   select parents from $X_1, \ldots, X_{i-1}$ such that
   $\mu(X_i|\text{Parents}(X_i)) = \mu(X_i|X_1, \ldots, X_{i-1})$
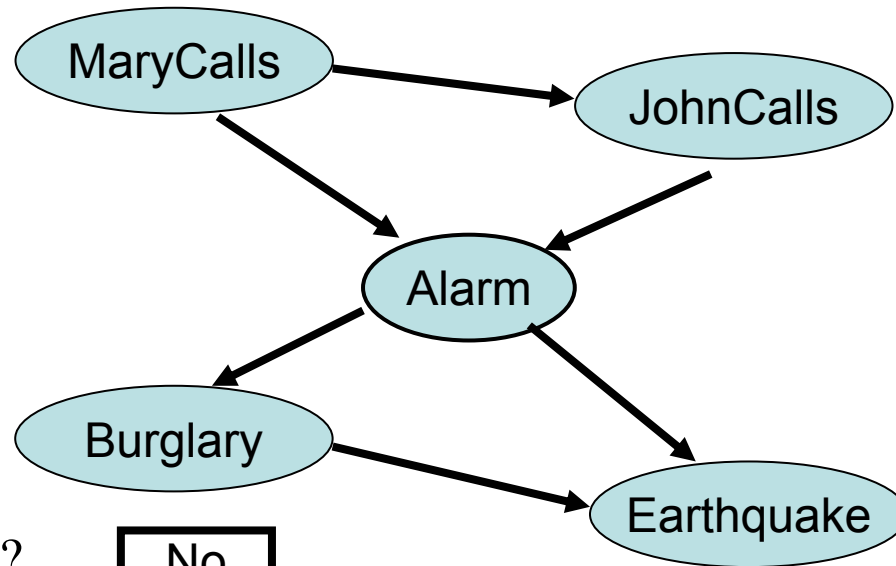
This choice guarantees the global semantics:
$$\mu(X_1, \ldots, X_n) = \Pi^n_{i=1}\, \mu(X_i|X_1, \ldots, X_{i-1}) \ \text{(chain rule)}$$
$$= \Pi^n_{i=1}\, \mu(X_i|\text{Parents}(X_i)) \text{ by construction}$$

Earthquake example with canonical ordering

- What is an appropriate ordering?

- In principle, each ordering is allowed!

- heuristic rule: start with causes, go to direct effects

- (B, E), A, (J, M) [4 possible orderings]

Earthquake example with noncanonical ordering

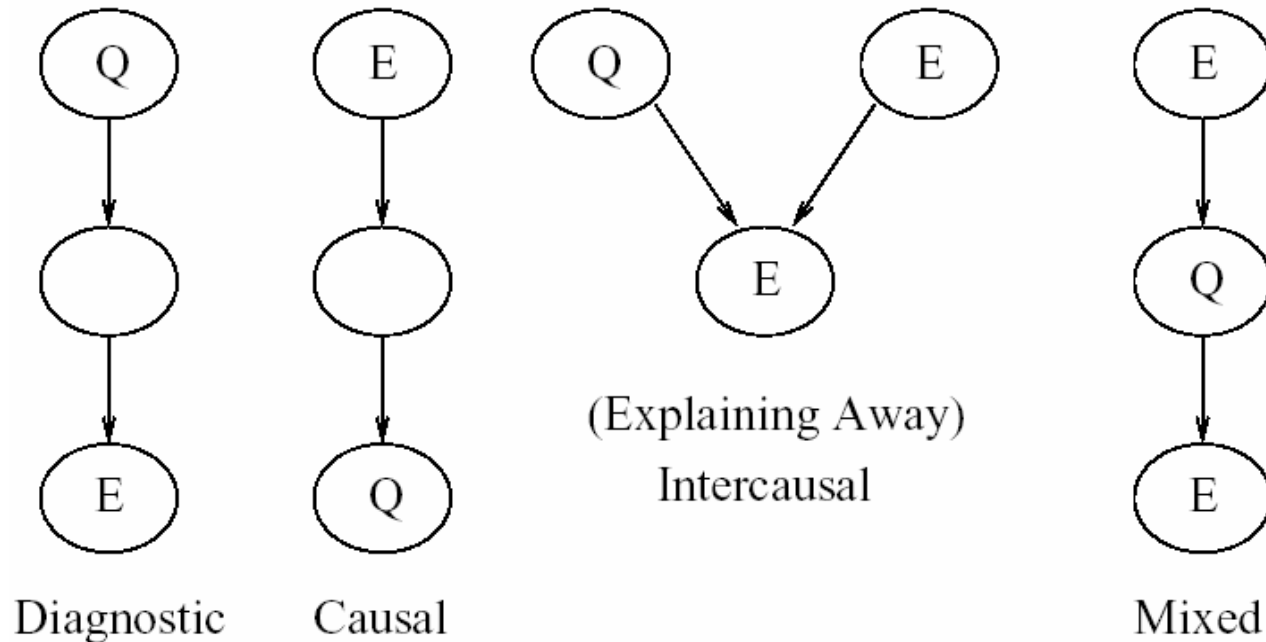- Suppose we chose the ordering M, J, A, B, E

- $\mu(J|M) = \mu(J)$? No
- $\mu(A|J,M) = \mu(A|J)$? $\mu(A|J,M) = \mu(A)$? No
- $\mu(B|A,J,M) = \mu(B|A)$? Yes
- $\mu(B|A,J,M) = \mu(B)$? No
- $\mu(E|B, A,J,M) = \mu(E|A)$? No
- $\mu(E|B,A,J,M) = \mu(E|A,B)$? Yes

6 Inference in belief networks

Types of inference:

Q quary variable, E evidence variable

Diagnostic   Causal   (Explaining Away) Intercausal   Mixed

Kinds of inference

- Diagnostic inferences: from effect to causes.

  `P(Burglary|JohnCalls)`

- Causal Inferences: from causes to effects.

  `P(JohnCalls|Burglary)`

  `P(MaryCalls|Burglary)`

- Intercausal Inferences: between causes of a common effect.

  `P(Burglary|Alarm)`

  `P(Burglary|Alarm ∧ Earthquake)`