Posts and Telecommunication Institute of Technology
Faculty of Information Technology 1

Introduction to Artificial Intelligence

Bayesian network

Ngo Xuan Bach
Dao Thi Thuy Quynh

Contents

- Definition and construction of Bayesian network
- Inference in Bayesian networks

Representing probability problem

- Inference problem:
  - Give evidences: $E_1, E_2, \dots, E_n$
  - Need to define requirement Q by computing $P(Q|E_1, E_2, \dots, E_n)$
- If there are all simultaneous probabilities
  - Can compute conditional probability above
- Simultaneous probability table has size that increases exponentially by the number of variables
  - Too big in reality

Need more realistic representation and inference methods

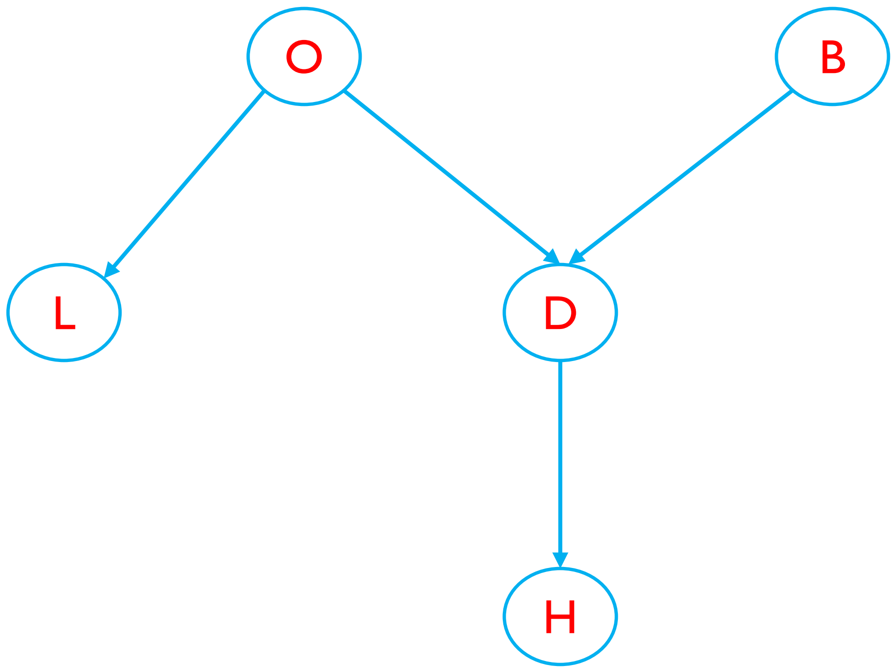
Example (1/2)

- Problem: A person comes home from work, need to guess if there is someone in the house?
- Know that:
  - If family members get away, the yard lights are often (not always) turned on
  - When there is no one at home, a dog is tied outside
  - If being sick, the dog is also tied outside
  - If the dog is outside, family members can hear the barking

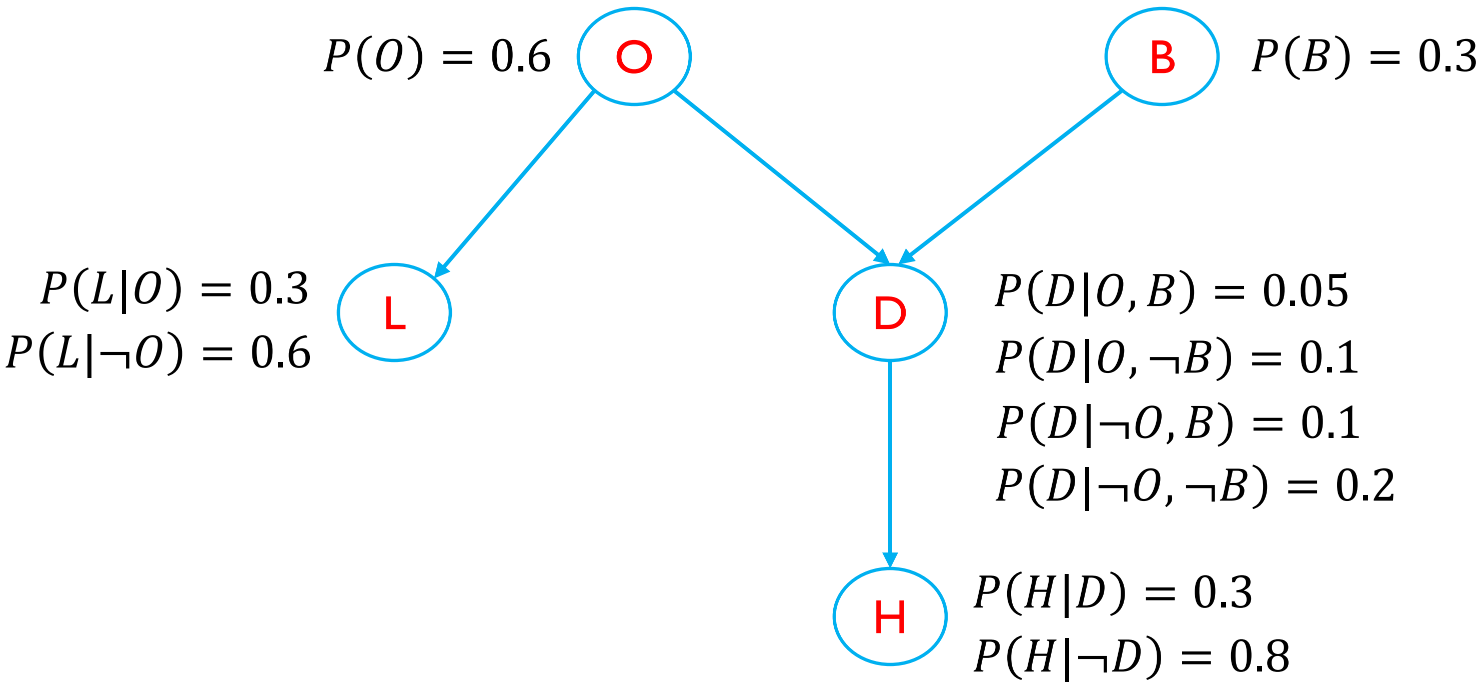
Example (2/2)

- Define the following 5 random variables:
  - O : there is no one at home
  - L : yard lights is turned on
  - D : the dog is tied outside
  - B : the dog is sick
  - H : can hear the barking

The relations between nodes



Bayesian network

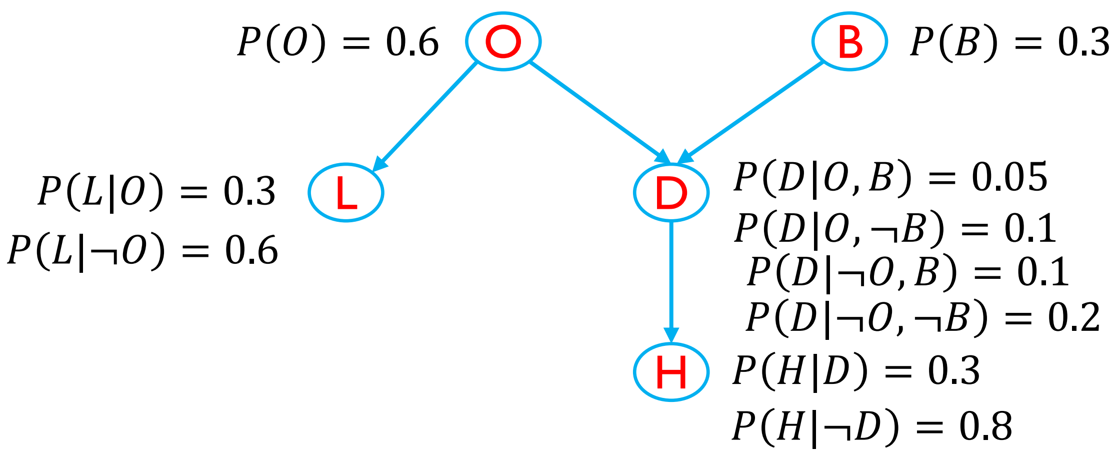


Definition of Bayesian network

- A Bayesian network includes 2 parts:
  - The first part is a directed acyclic graph, in which each node corresponds to a random variable, and each (directed) edge represents the relation between the root node and the targeted node.
  - The second part is a conditional probability table, including a conditional probability of each child node given the combinations of values of its parent nodes.

Independence probability in Bayes Network

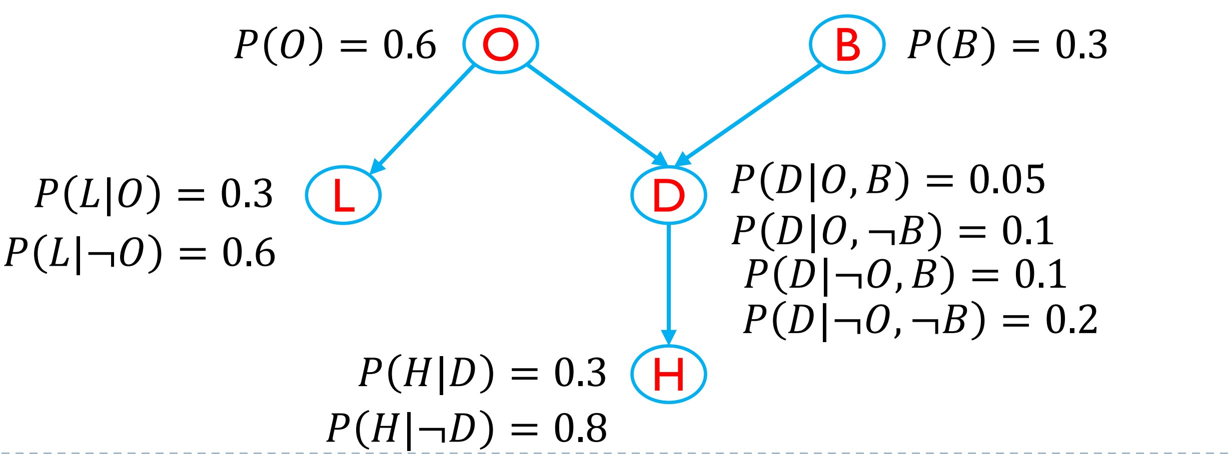
- Bayesian networks allow to represent briefly all simultaneous probabilities
  - The reduction can be done by using the probability independence feature in the network.
- Independence probability
  - Each node V is independent of all nodes that are not descendants of V, given the values of parent nodes of V.
  - Example: H is conditional independent of L, O, B if knowing D



Compute simultaneous probability in Bayesian Network

- Example:

$$
\begin{aligned}
&P(H, \neg L, D, \neg O, B) \\
&\quad = P(H \mid \neg L, D, \neg O, B)\, P(\neg L, D, \neg O, B) \\
&\quad = P(H|D)\, P(\neg L, D, \neg O, B) \\
&\quad = P(H|D)\, P(\neg L \mid D, \neg O, B)\, P(D, \neg O, B) \\
&\quad = P(H|D)\, P(\neg L|\neg O)\, P(D, \neg O, B) \\
&\quad = P(H|D)\, P(\neg L|\neg O)\, P(D \mid \neg O, B)\, P(\neg O, B) \\
&\quad = P(H|D)\, P(\neg L|\neg O)\, P(D|\neg O, B)\, P(\neg O)\, P(B) \\
&\quad = (0.3)(1 - 0.6)(0.1)(1 - 0.6)(0.3)
\end{aligned}
$$



Compute simultaneous probability in Bayesian Network (General)

$$P(X_1 = x_1, \ldots, X_n = x_n) = \prod_{i=1}^{n} P(X_i = x_i \mid parents(X_i))$$

Bayesian Network Construction

- Two ways to build Bayesian networks:
  - By hand (by human)
    - Base on the knowledge of human about the problem
    - Include 2 steps: define the structure of graph and fill values in conditional probability table
  - Machine learning from data: in case there are data of combinations of variables
    - Probabilities distribution presented by network best fits the frequency of occurrence of values in the data set

Bayesian Network Construction (by hand)

1. Define of the set of related random variables
2. Choose the order of variables

   Example: $X_1, X_2, \dots, X_n$

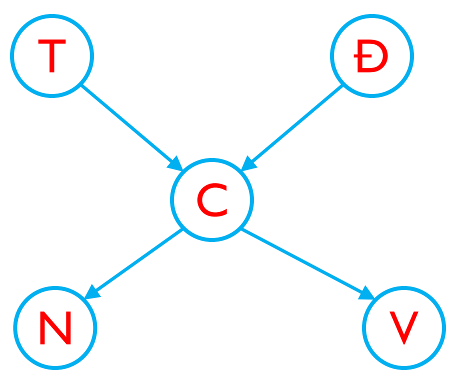
3. **for** $i = 1$ **to** n **do**
   a. Add a node for X_i
   b. Select $parents(X_i)$ is the smallest set of existing nodes so that X_i is conditionally independent of all previous nodes given $parents(X_i)$
   c. Add a directed arc from each node $parents(X_i)$ to X_i
   d. Add conditional probability values $P(X_i | parents(X_i))$ or $P(X_i)$ if $parents(X_i) = \emptyset$

Example (1/2)

- A person installed an anti-theft alarm system at home.
- The system will alarm when there is a thief.
- But, the system can make a false alarm if there is a tremor by an earthquake.
- In case of hearing the alarm, two neighbors Nam and Việt will call the house onwer
- Due to many difference reasons, Nam and Việt can announce inaccurately, for instance, due to noise they can not hear any alarm or vice verser, They mistake other sounds for the alarm.

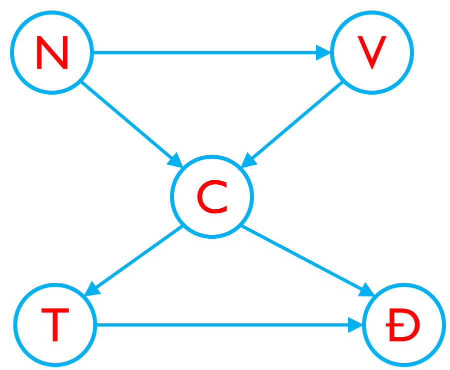
Example (2/2)

- **Step 1**: Select variables: use 5 following variables:
  - T (having thief), Đ (earthquake), C (the system alarm), N (Nam calls), V (Việt calls)
- **Step 2**: Variables are arranged by order: T, Đ, C, N, V
- **Step 3**: Following the steps in the figure, we can build the network (for simplicity, the figure only shows the structure and does not have the conditional probability table).



The impact of ordering nodes

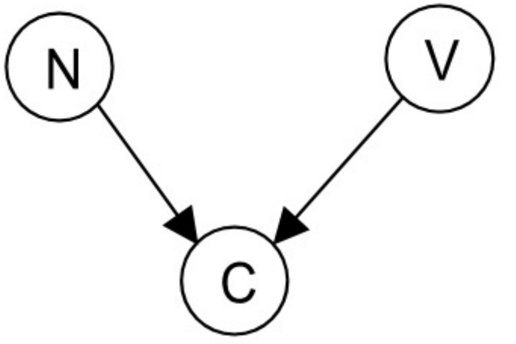
- Building a Bayesian network in reality is not simple
  - Choosing the right order of nodes, from which we can select the set of father nodes with small size, is difficult.
- Assume that variables are arranged by a different order: $N, V, C, T,$ Đ



General probability independent feature: Definition of d-seperation (1/5)

- If the value of node C is not given:
  - According to the independent feature of Bayesian Network, N and V are independent (unconditional)
- If the value of node C is given
  - Are N and V still independent?

The learned knowledge does not allow to answer this question



General probability independent feature: Definition of d-seperation (2/5)

- *d-seperation* answers the question about the independence of a set of nodes X with a set of nodes Y given the set of nodes E on a Bayesian network.
  - Nodes X and nodes Y are called as being *d-seperated* by nodes E if X and Y are independent given E.
  - Nodes X and nodes Y are *d-connection* if they are not *d-seperated*
- To define the *d-seperation* of sets X and Y, we first define *d-seperation* between 2 single nodes x of X and y of Y
  - 2 sets of nodes will be independent if each node in one set is independent of all nodes in the other.

General probability independent feature: Definition of d-seperation(3/5)

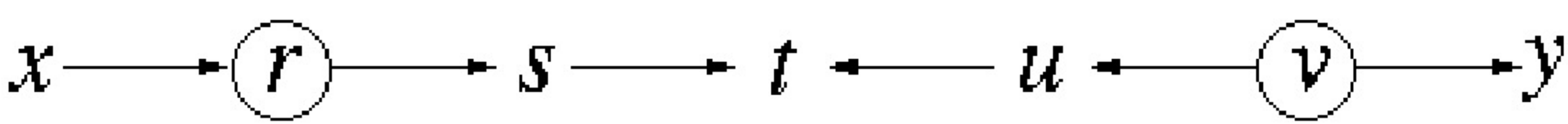
- **Principle 1**: Node x and y are d*-connected* if there is an unblocked path *between 2 nodes*. In contrast, if there is no such path, x and y are d*-seperated.*
  - A path is a sequence of contiguous arcs, regardless of the direction of the arcs.
  - An unblocked path is a path on which no 2 adjacent arcs are directed at each other.
  - A node with two such inward arcs is called as a conflict node

$$x \longrightarrow r \longrightarrow s \longrightarrow t \longleftarrow u \longleftarrow v \longrightarrow y$$

  - The connection and seperation features following **Principle 1** is unconditional and so the independence of probability is defined by **Principle 1** is unconditionally independent.

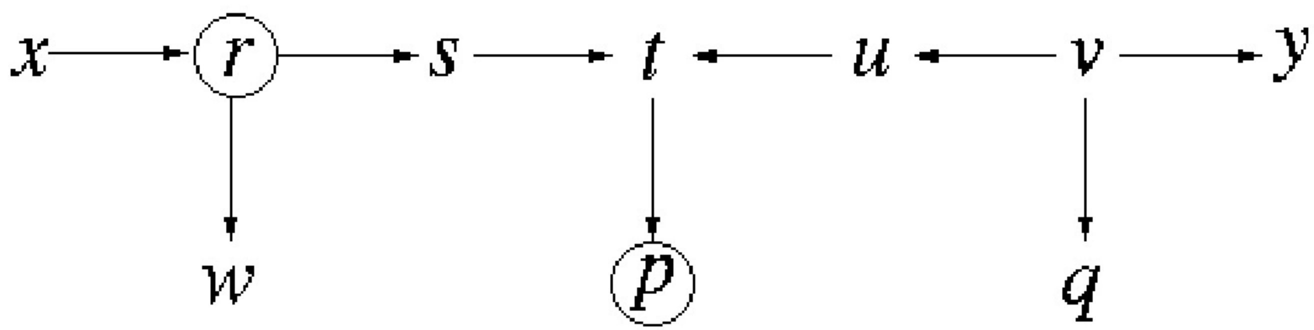
General probability independent feature: Definition of d-seperation(4/5)

- **Principle 2**: node x and y are *d-conditional connected* given the set of nodes E if there exists an unblocked path (not include any confict nodes) and does not pass any nodes of E. In contrast, if there is no such path, we say that x and y are *d-seperated* by E. In other words, every paths between x and y (if any) are blocked by E.
  - When knowing the value of some nodes (set of nodes E), the independence or dependence between remanining nodes can be changed
  - the independence or dependence in this case is called as *d-conditional seperation* by the set of variables E



General probability independent feature: Definition of d-seperation (5/5)

- **Principle 3**: If a conflict node is member of set E, or having descendant in set E, so that node does not block paths through it
  - Assume that we know that an event is caused by 2 or more causes, if we already know 1 cause is true then the probability of the other causes is reduced, if we know 1 cause is false then the probability of other causes increases



s and y are *d-connection*, x and u are *d-seperation*

Contents

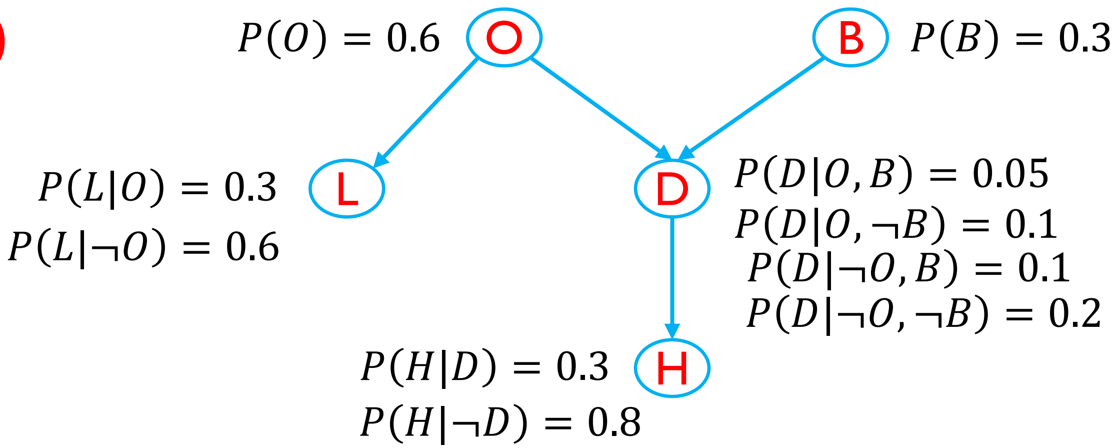
- Definition and construction of Bayesian network
- Inference in Bayesian networks

Lessons learned (recall)

- How to build a bayes network (by hand)
- Bayesian networks allow to reduce the representation
  - No need to save the entire simultaneous probability table
- Can compute simultaneous probability of any combination of values of variables
- Therefore, to compute every posterior probability needed for inference

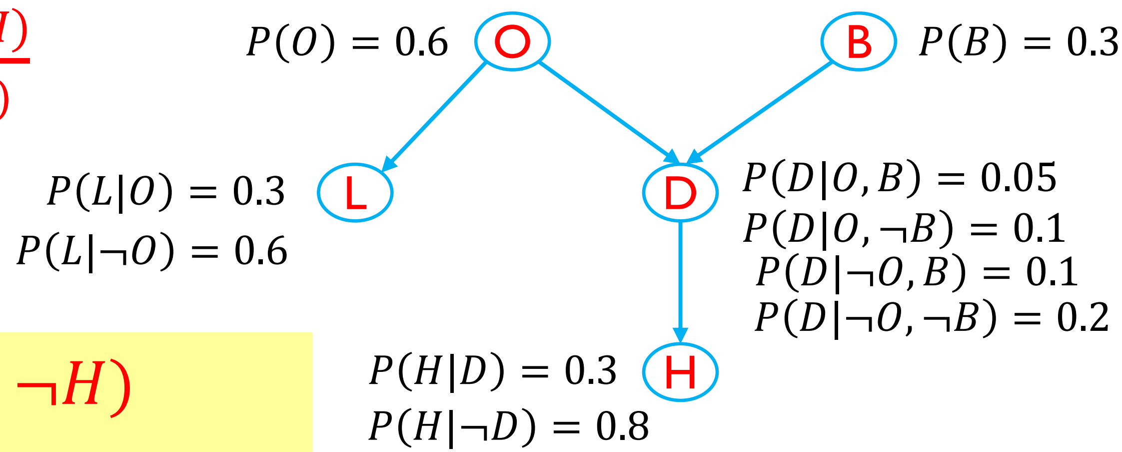
Example of posterior probability

- Compute $P(L \mid B, \neg H)$



Example of posterior probability

- $P(L|B,\neg H) = \frac{P(L,B,\neg H)}{P(B,\neg H)}$



Step 1: compute $P(L,B,\neg H)$

Step 2: compute $P(\neg L,B,\neg H)$

Step 3: compute
$$\frac{P(L,B,\neg H)}{P(L,B,\neg H)+P(\neg L,B,\neg H)}$$

Simultaneous probability can be computed as (in) the previous lesson.

General case

$$P(E_1|E_2) = \frac{P(E_1 \wedge E_2)}{P(E_2)}$$

$$= \frac{The\ sum\ of\ simultaneous\ probabilities\ including\ E_1 and\ E_2}{The\ sum\ of\ simultaneous\ probabilities\ including E_2}$$

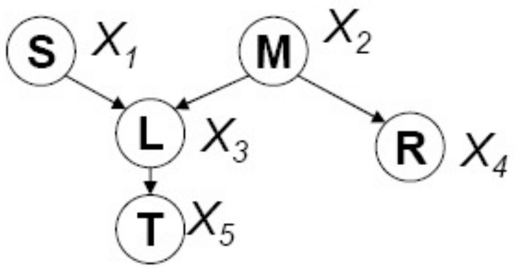
- Problem
  - Require to list simultaneous probabilities having E_1, E_2
  - The number of simultaneous probabilities increases exponentially by the number of variables ⇒ unrealistic

- **Inference in general in Bayesian networks is an NP-hard problem ☹ ☹ ☹**

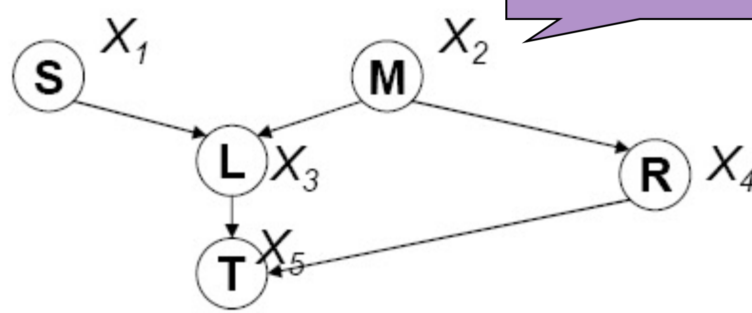
Inference in reality

- Inference for a particular case
  - When network has the form of single connection (poly tree): there is no more than 1 path between any 2 nodes

NOT



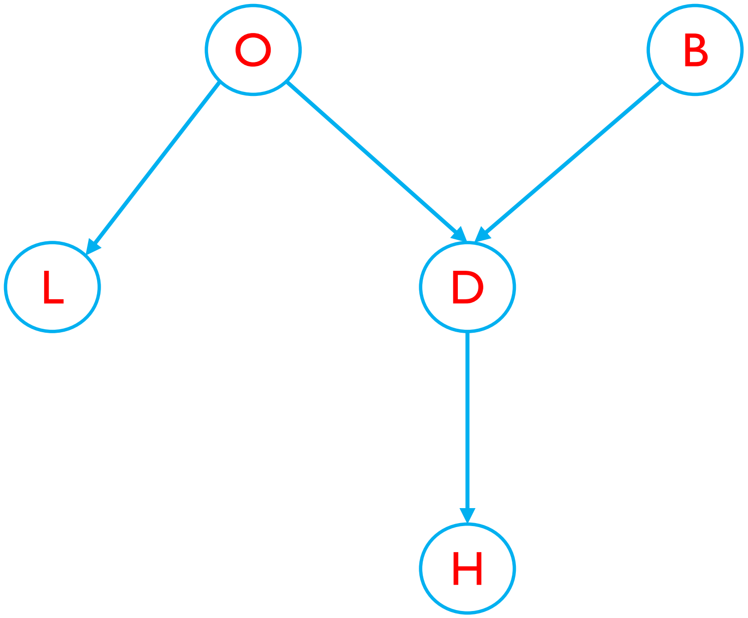
A poly tree



  - There exists an algorithm with linear complexity for poly trees.
- Approximate inference by sampling

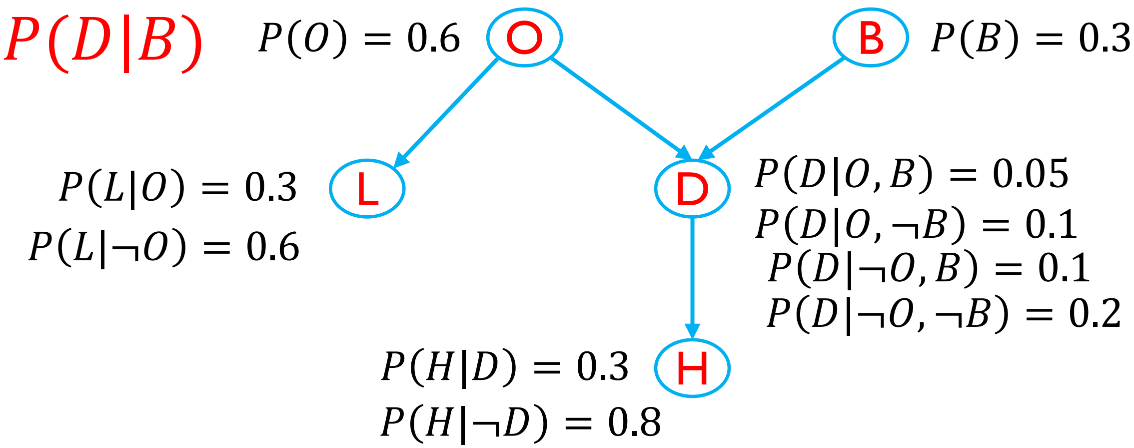
Inference for a particular case

- The simplest case:
  - When the evidence E and result Q have only one direct connection
  - Distinguish 2 cases:
    - Causal inference (top to bottom): need to compute $P(Q|E)$ when E is the parent node of Q
    - Diagnostic inference (bottom to top): need to compute $P(Q|E)$ when E is the child node of Q



Causal inference (1/3)

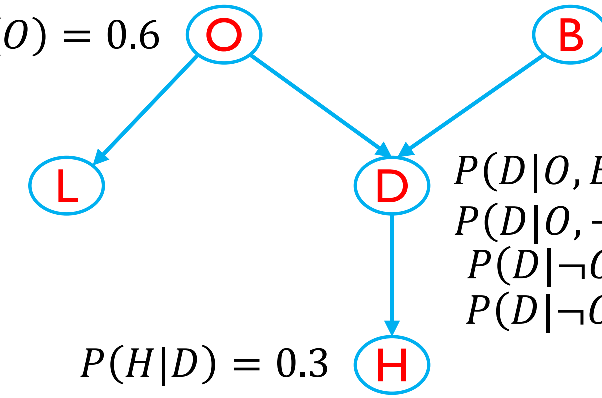
- Example: compute $P(D|B)$



Causal inference (2/3)

- Example: compute $P(D|B)$

$P(O) = 0.6$ $\quad P(B) = 0.3$



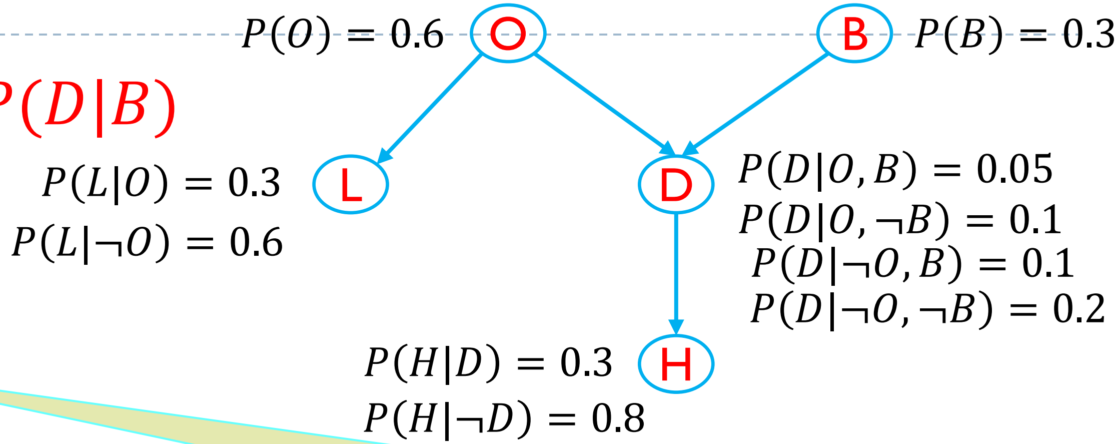
$P(L|O) = 0.3$
$P(L|\neg O) = 0.6$

$P(D|O,B) = 0.05$
$P(D|O,\neg B) = 0.1$
$P(D|\neg O,B) = 0.1$
$P(D|\neg O,\neg B) = 0.2$

$P(H|D) = 0.3$
$P(H|\neg D) = 0.8$

$$P(D|B) = \frac{P(D,B)}{P(B)}$$
$$= \frac{P(D,B,O) + P(D,B,\neg O)}{P(B)}$$
$$= \frac{P(D|B,O)P(B,O) + P(D|\, B,\neg O)P(B,\neg O)}{P(B)}$$
$$= \frac{P(D|B,O)P(B)P(O) + P(D|\, B,\neg O)P(B)P(\neg O)}{P(B)}$$
$$= P(D|B,O)P(O) + P(D|\, B,\neg O)P(\neg O)$$
$$= (0.05)(0.6) + \ (0.1)(1 - 0.6)$$
$$= 0.07$$

Diagnostic inference (3/3)



- Example: compute $P(D|B)$

$$P(D|B) = \frac{P(D,B)}{P(B)}$$

$$= \frac{P(D,B,O) + P(D,B,\neg O)}{P(B)}$$

Step 1: Convert conditional probability to simultaneous probability

$$= \frac{P(D|B,O)P(B,O)+P(D|\,B,\neg O)P(B,\neg O)}{P(B)}$$

$$= \frac{P(D|B,O)P(B)P(O)+P(D|\,B,\neg O)P(B)P(\neg O)}{P(B)}$$

$$= P(D|B,O)P(O) + P(D|\,B,\neg O)P(\neg O)$$

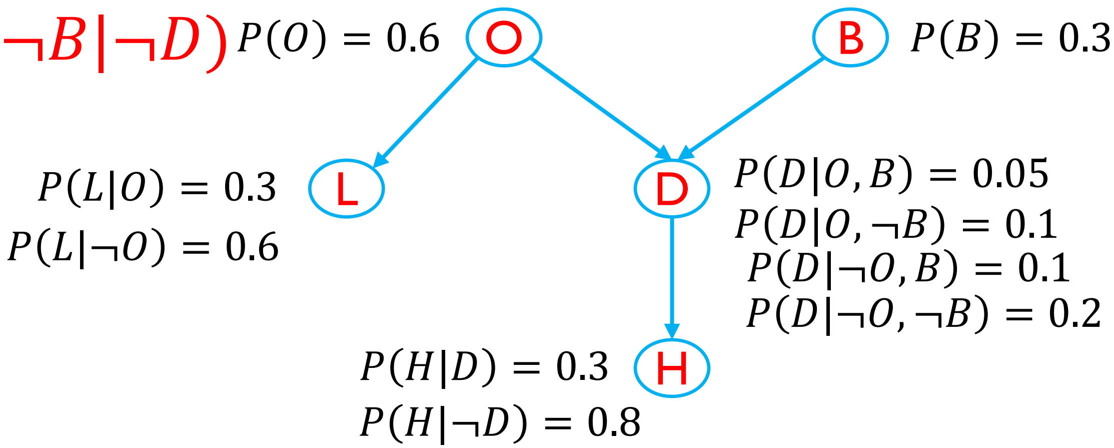
Step 2: Use the independent feature of probability in Bayesian networks, rewrite simultaneous probability in form conditional probabilities of child node when knowing values of parent nodes

$$= (0.05)(0.6) + \ (0.1)(1 - 0.6)$$
$$= 0.07$$

Step 3: Use probability values from conditional probability table to compute

Diagnostic inference (1/5)

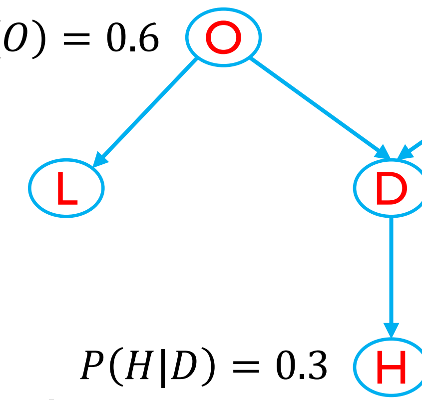
- Example: compute $P(\neg B | \neg D)$



Diagnostic inference(2/5)

- Follow Bayes

$$P(\neg B|\neg D) = \frac{P(\neg D|\neg B)P(\neg B)}{P(\neg D)}$$

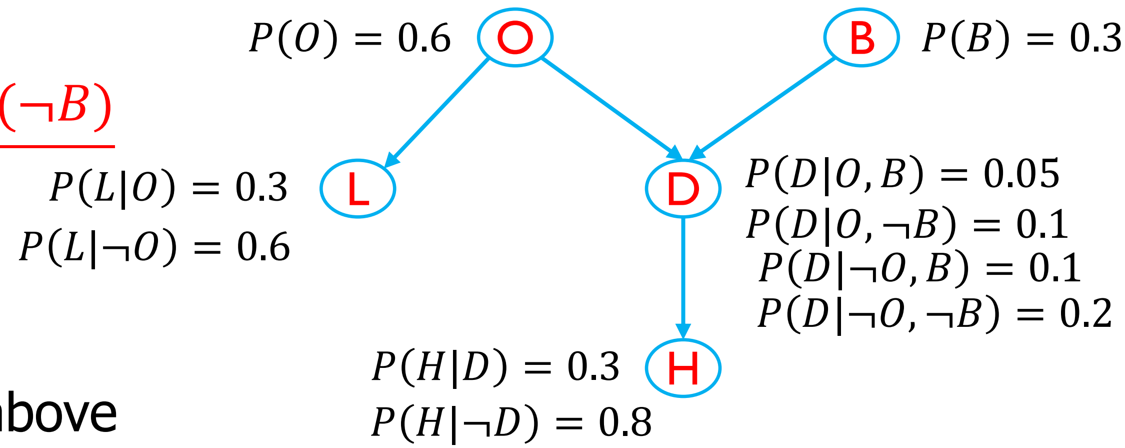


compute $P(\neg D|\neg B)$ like the previous part

Diagnostic inference(3/5)

- Follow Bayes

$$P(\neg B|\neg D) = \frac{P(\neg D|\neg B)P(\neg B)}{P(\neg D)}$$



compute $P(\neg D|\neg B)$ as above

$$\begin{aligned} P(\neg D|\neg B) &= P(\neg D|O,\neg B)P(O) + P(\neg D|\neg O,\neg B)P(\neg O) \\ &= (0.9)(0.6) + (0.8)(0.4) \\ &= 0.86 \end{aligned}$$

$$P(\neg B|\neg D) = \frac{(0.86)(0.7)}{P(\neg D)} = \frac{0.602}{P(\neg D)}$$

To compute $P(\neg D)$, we will compute $P(B|\neg D)$

Diagnostic inference (4/5)

- $P(B|\neg D) = \frac{P(\neg D|B)P(B)}{P(\neg D)} = \frac{(1-0.07)0.3}{P(\neg D)} = \frac{0.279}{P(\neg D)}$
- Use

$P(\neg B|\neg D) + P(B|\neg D) = 1$

$$\frac{0.602}{P(\neg D)} + \frac{0.279}{P(\neg D)} = 1$$

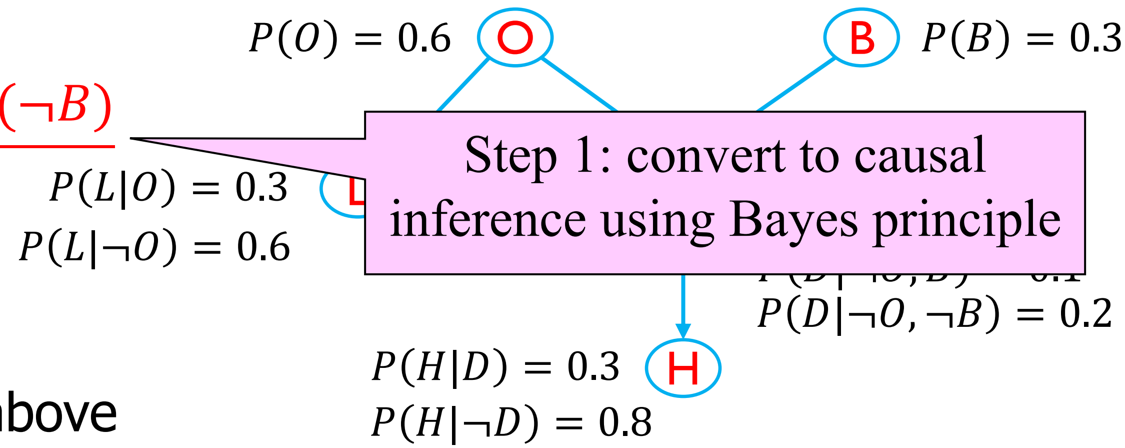
Then $P(\neg D) = 0.881$

Replace:

$$P(\neg B|\neg D) = \frac{0.602}{P(\neg D)} = \frac{0.602}{0.881} = 0.683$$

- Follow Bayes

$$P(\neg B|\neg D) = \frac{P(\neg D|\neg B)P(\neg B)}{P(\neg D)}$$



Step 1: convert to causal inference using Bayes principle

compute $P(\neg D|\neg B)$ as above

$$P(\neg D|\neg B) = P(\neg D|O,\neg B)P(O) + P(\neg D|\neg O,\neg B)P(\neg O)$$
$$= (0.9)(0.6) + (0.8)(0.4)$$
$$= 0.86$$

Step 2: perform the same as causal inference

$$P(\neg B|\neg D) = \frac{(0.86)(0.7)}{P(\neg D)} = \frac{0.602}{P(\neg D)}$$

To compute $P(\neg D)$, we will compute $P(B|\neg D)$

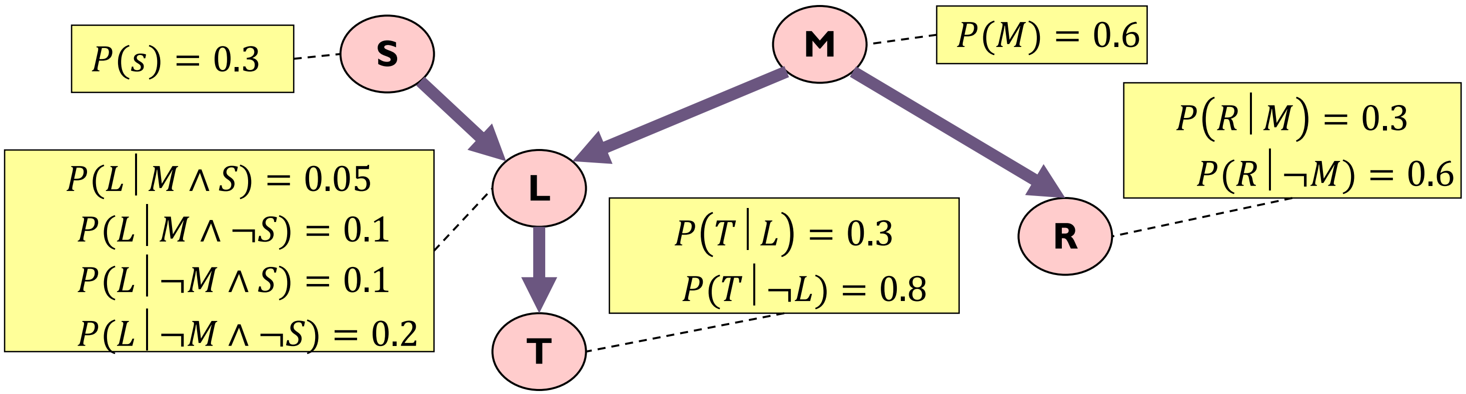
General method

- Applies to both causal inference and diagnostic inference
  - **Step 1**: Convert conditional probability to simultaneous probability
  - **Step 2**: Use the independent feature of probability in Bayesian networks, rewrite simultaneous probability in form conditional probabilities of child node when knowing values of parent nodes
  - **Step 3**: Use probability values from the conditional probability table to compute

Inference by sampling

- In general case: Inference in Bayesian networks is an NP-hard problem (very complicated).
- Can infer approximately by sampling
- Generate sets of variables with the same simultaneous probabilities of the network

Sampling(1/2)



- Ramdomly select S: $S = true$ with probability 0.3
- Ramdomly select M: $M = true$ with probability 0.6
- Ramdomly select L: probability $L = true$ depends on the values of S, M above
  - Assume that above $steps\ generate\ M = true, S = false$, $L = true$ with probability 0.1
- Ramdomly select R with probability depend on value of M
- Ramdomly select T with probability depend on value of L

Sampling(2/2)

- Assume the we need to compute: $P(R = True \mid T = True, S = False)$
- Sampling many times as above, each generated set of values is called a sample
- Compute the number of occurrence of events:
  - Nc: number of samples having $T = True$ and $S = False$
  - Ns: number of samples having $R = True$, $T = True$ and $S = False$
  - N: Total number of samples
- If N is large enough:
  - Nc/N: (approximately) probability $P(T = True\ and\ S = False)$
  - Ns/N: (approximately) probability $P(R = True\ , T = True\ , S = False\ )$
  - $P(R|T, \neg S) = P(R, T, \neg S)/P(T, \neg S) \approx Ns/Nc$

Generally sampling

- Need to compute $P(E_1|E_2)$
- Sample large enough quantity
- Compute quantity:
  - Nc: number of samples having E_2
  - Ns: number of samples having E_1 and E_2
  - N: Total number of samples
- If N is large enough, we have: $P(E_1|E_2) = \frac{Ns}{Nc}$

Exercise

- Do some exercises in the textbook