SNS COLLEGE OF ENGINEERING

Kurumbapalayam(Po), Coimbatore – 641 107
Accredited by NAAC-UGC with 'A' Grade
Approved by AICTE, Recognized by UGC & Affiliated to Anna University, Chennai

Department of AI &DS

Course Name – 23ADT201 ARTIFICIAL INTELLIGENCE

II Year / III Semester

UNIT 5

PROBABILISTIC REASONING

BAYESIAN NETWORK

CASE STUDY:

A case study on Bayesian networks in AI is medical diagnosis systems, where they model relationships between symptoms and diseases. By using probability to update beliefs based on observed symptoms, they assist in diagnosing conditions with uncertain data.

Joint probability distribution

- The **full joint probability distribution** specifies the probability of values to random variables.
- It is usually too large to create or use in its explicit form.
- Joint probability distribution of two variables X and Y are

Joint probabilities	X	X'
Y	0.20	0.12
Y'	0.65	0.03

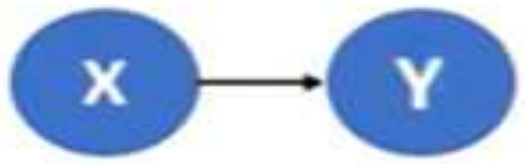
- Joint probability distribution for n variables require 2^n entries with all possible combination.

Drawbacks of joint Probability Distribution

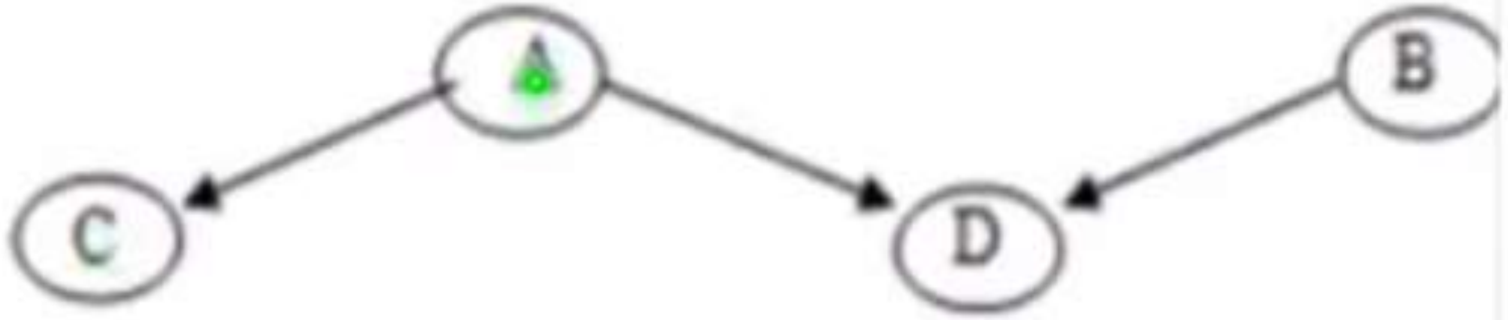
- Large number of variables and grows rapidly
- Time and space complexity are huge
- Statistical estimation with probability is difficult
- Human tends signal out few propositions
- The alternative to this is Bayesian Networks.

Bayesian Networks

- Bayesian Network is to represent the dependencies among variables and to give a brief specification of any full joint probability distribution.
- A Bayesian network is a directed graph in which each nodes are variables and edges are relations.
- The full specification is as follows:
  - 1. A set of random variables makes up the nodes of the network. Variables may be discrete or continuous.
  - 2. A set of directed links or arrows connects pairs of nodes. If there is an arrow from node *X* to node *Y*, X is said to be a parent of *Y*.
  - 3. Each node *X*, has a conditional probability distribution *P(X,(Parents(X,))* that quantifies the effect of the parents on the node. (X is parent of Y)
  - 4. The graph has no directed cycles (and hence is a directed, acyclic graph, or DAG).



Example



Bayesian Belief Network

- A & B are unconditional, independent, evidence and parent nodes
- C & D are conditional, dependent, hypothesis and child nodes.

Conditional Probability Table

P(A)	=	0.3
P(B)	=	0.6
P(C|A)	=	0.4
P(C|~A)	=	0.2
P(D|A, B)	=	0.7
P(D|A, ~B)	=	0.4
P(D|~A, B)	=	0.2
P(D|~A, ~B)	=	0.01

Conditional Probability Tables

P(A)	P(B)
0.3	0.6

A	P(C)
T	0.4
F	0.2

A	B	P(D)
T	T	0.7
T	F	0.4
F	T	0.2
F	F	0.01

- P(A,B,C,D) = P(D|A,B)*P(C|A)*P(B)*P(A)

Example 2

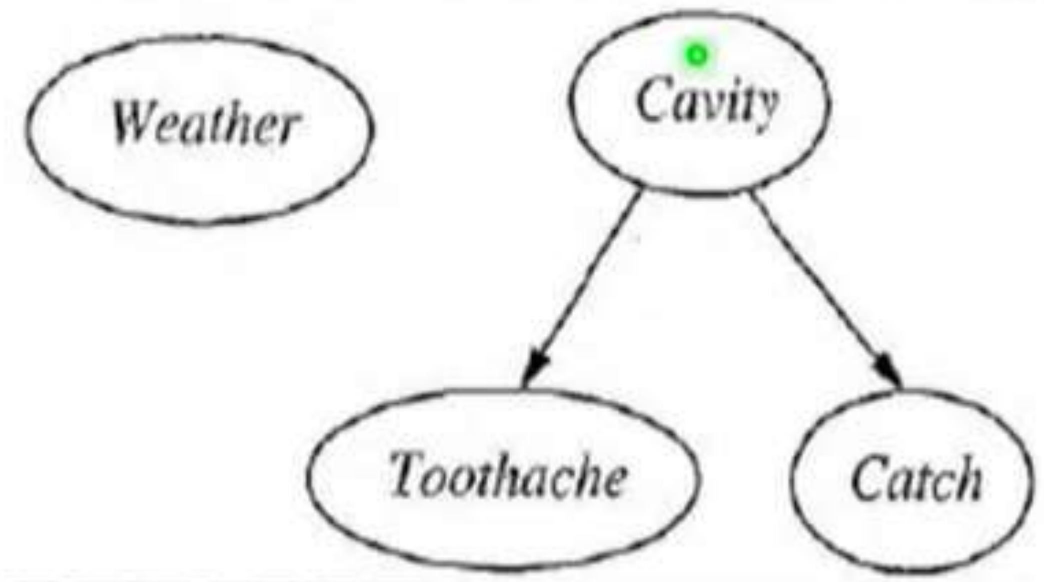


Figure 14.1 A simple Bayesian network in which *Weather* is independent of the other three variables and *Toothache* and *Catch* are conditionally independent, given Cavity.

Example -3 - Burglar alarm

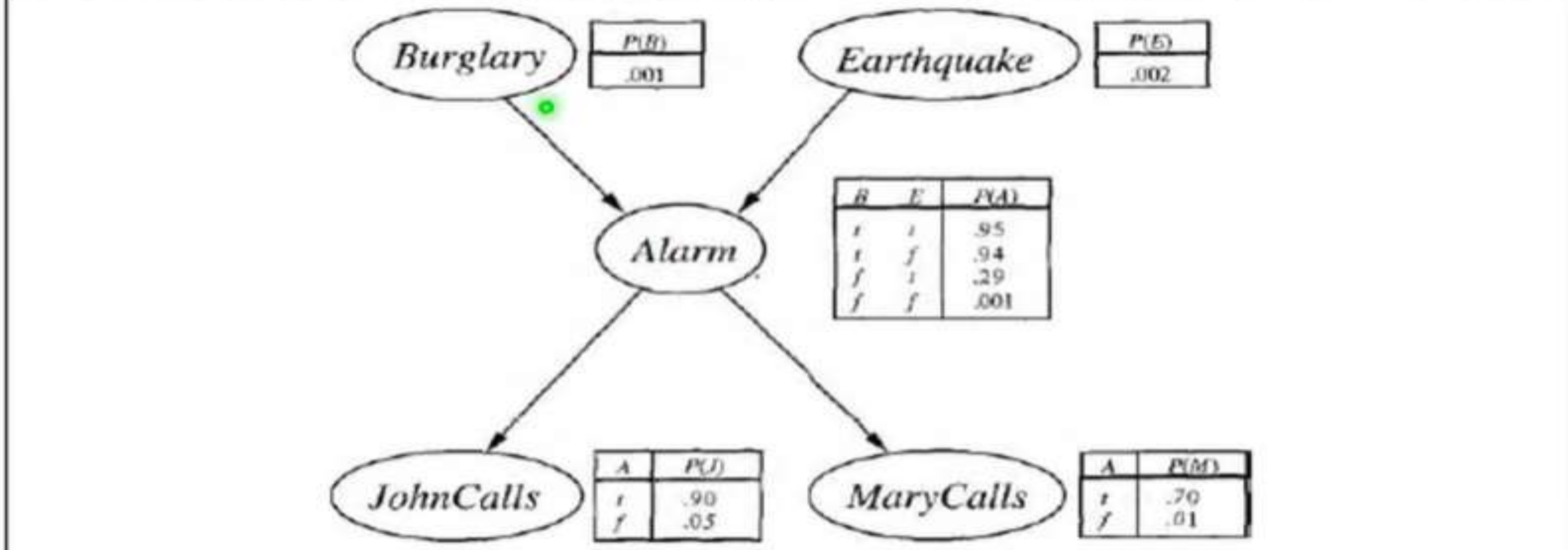
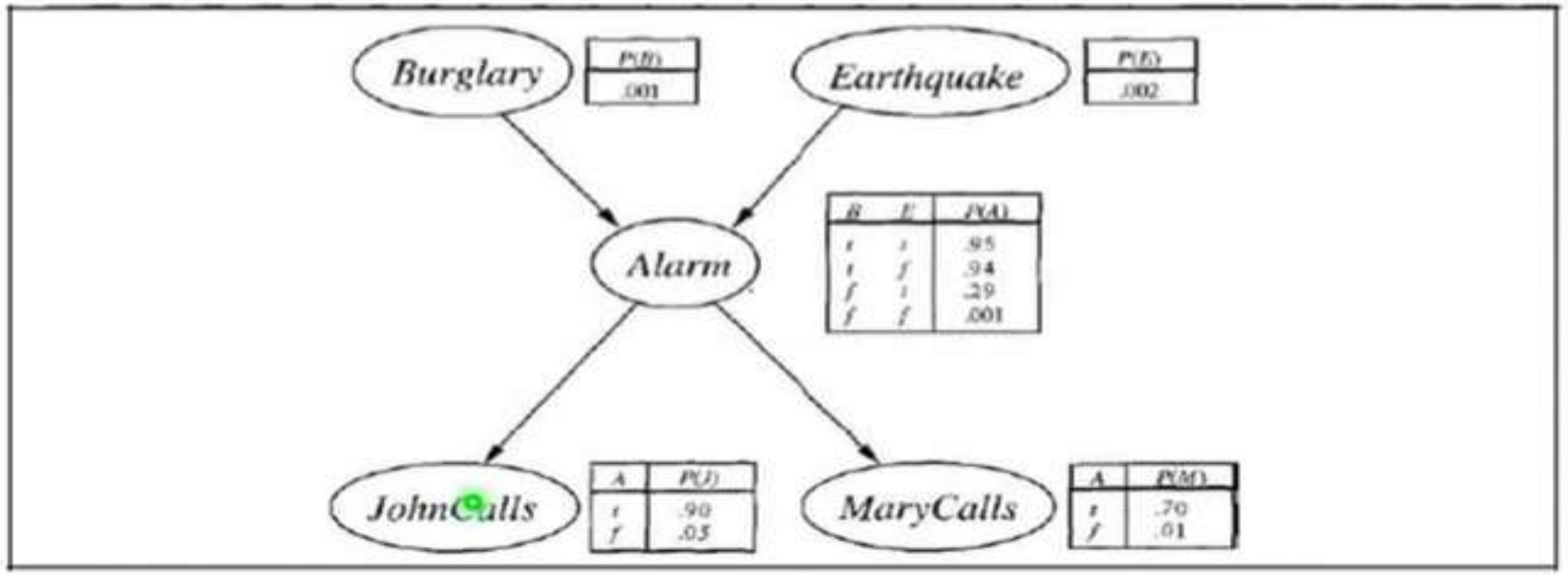


Figure 14.2 A typical Bayesian network, showing both the topology and the conditional probability tables (CPTs). In the CPTs, the letters B, E, A, J, and M stand for *Burglary*, *Earthquake*, *Alarm*, *JohnCalls*, and *MaryCalls*, respectively.

Example - Burglar alarm

- You have a new burglar alarm installed at home.
- It also responds on occasion to minor earthquakes.
- You also have two neighbors, John and Mary, they promised to call you when they hear the alarm.
- John always calls when he hears the alarm, but sometimes confuses the telephone ringing with the alarm, and calls then, too.
- Mary likes rather loud music, and sometimes, she misses the alarm altogether.
- Given the evidence of who has or has not called, we would like to estimate the probability of a burglary.

- The burglary and earthquakes directly affect the probability of the alarm's going off,
- But, John and Mary call depends only on the alarm.
- The network does not have nodes for Mary's currently listening to loud music or the telephone ringing and confusing John.



Example

- We can calculate the probability that the alarm has sounded, but neither a burglary nor an earthquake has occurred, and both John and Mary call.

$$P(j \wedge m \wedge a \wedge \neg b \wedge \neg e)$$
$$= P(j|a)P(m|a)P(a|\neg b \wedge \neg e)P(\neg b)P(\neg e)$$
$$= 0.90 \times 0.70 \times 0.001 \times 0.999 \times 0.998 = 0.00062.$$

Semantics of Bayesian Network

- An entry in joint distribution is the probability of **conjunction** of particular assignment to each variable, such as
- $P(X1=x_1 \wedge X2=x_2 \wedge \ldots \wedge .. Xn=x_n)$ is equal to
- $P(x_1,x_2,\ldots,x_n) = \prod_{i=1}^{n} P(x_i \mid Parent(X_i))$

Method for Constructing Bayesian Network

- Rewrite the joint distribution in terms of a conditional probability, using the **product rule**

$$P(x_1, \ldots, x_n) = P(x_n | x_{n-1}, \ldots, x_1) P(x_{n-1}, \ldots, x_1)$$

- Then we repeat the process, reducing each conjunctive probability to a conditional probability and a smaller conjunction. We end up with one big product:

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

$$\mathbf{P}(X_i | X_{i-1}, \ldots, X_1) = \mathbf{P}(X_i | Parents(X_i)) ,$$

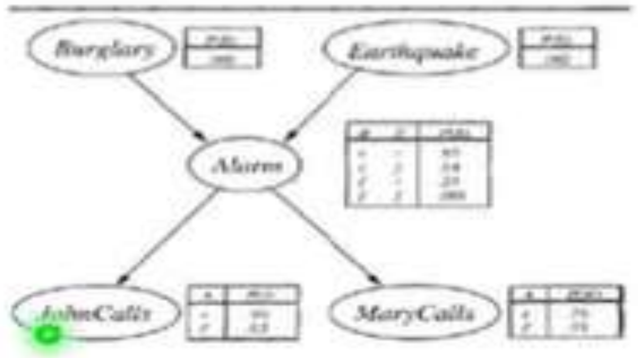
BAYESIAN NETWORK

$$\mathbf{P}(X_i | X_{i-1}, \ldots, X_1) = \mathbf{P}(X_i | Parents(X_i)) \,,$$

$$\mathbf{P}(MaryCalls | JohnCalls, Alarm, Earthquake, Burglary) = \mathbf{P}(MaryCalls | Alarm) \,.$$

Compactness and Node Ordering

- The compactness of Bayesian network is an example of general property of locally constructed systems. (also called as spare systems, inside some components there, and those are communicated)
- In a locally structured system, each subcomponent interacts directly with only a bounded number of other components, regardless of the total number of components.
- Therefore the correct order in which to add node is to add the 'root causes' first, then the variables they influenced and so on until we reach the leaves.



- Suppose we decide to add the nodes in the order MaryCalls, JohnCalls, Alarm, Burglary, Earthquake.
- Adding Mary Calls: No parents.
- Adding JohnCalls: If Mary calls, that probably means the alarm has gone off, which of course would make it more likely that John calls. Therefore, JohnCalls needs Mary Calls as a parent
- Adding Alarm: Clearly, if both call, it is more likely that the alarm has gone off than if just one or neither call, so we need both MaryCalls and JohnCalls as parents.
- Adding Burglary: If we know the alarm state, then the call from John or Mary might give us information about our phone ringing or Mary's music, but not about burglary:
- **P(Burglary|Alarm, JohnCalls, MaryCalls) = P(BurglaryIAlarm)**
- Hence we need just Alarm as parent.
- Adding *Earthquake:* if the alarm is on, it is more likely that there has been an earthquake. But if we know that there has been a burglary, then that explains the alarm, and the probability of an earthquake would be only slightly above normal. Hence, we need both *Alarm* and *Burglary* as parents.

BAYESIAN NETWORK

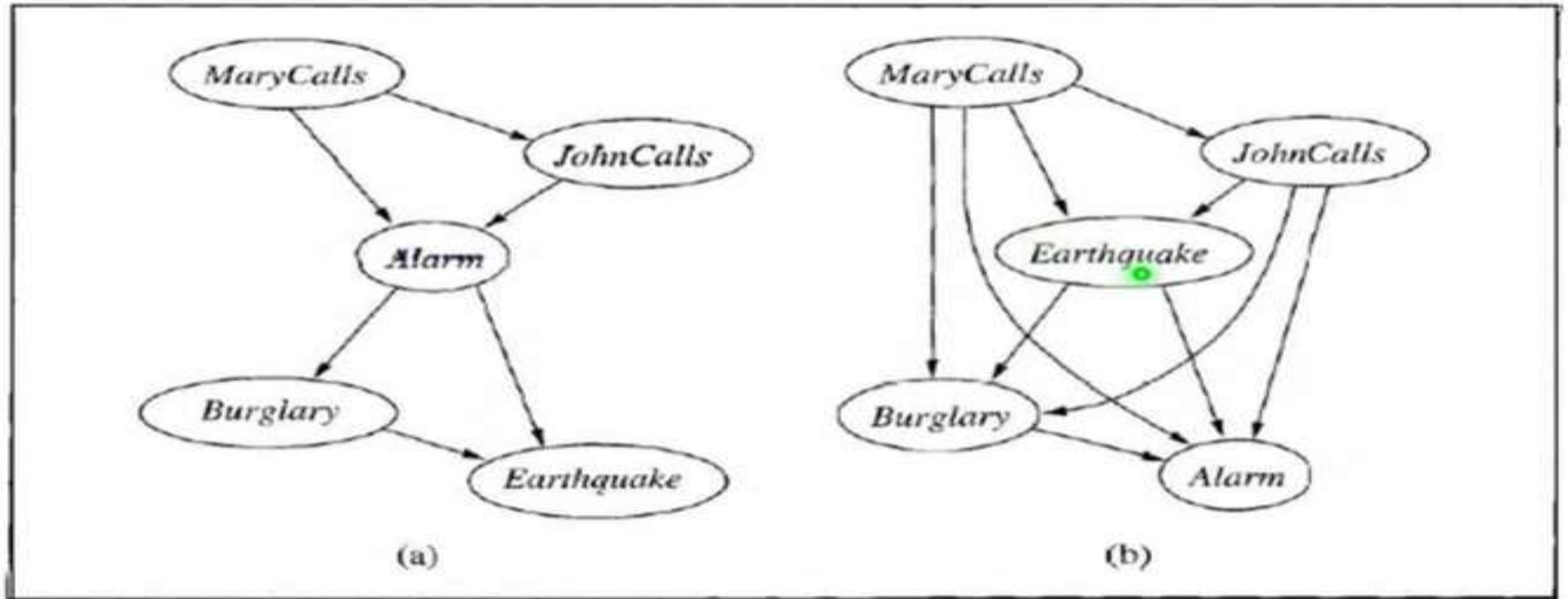
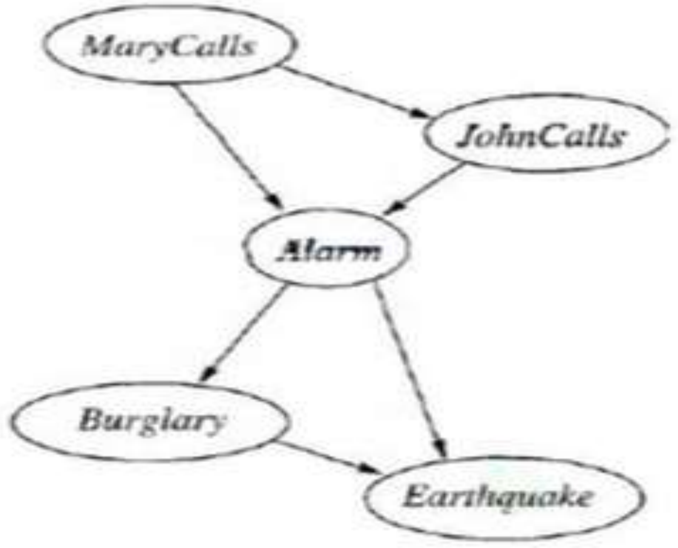


Figure 14.3 Network structure depends on order of introduction. In each network, we have introduced nodes in top-to-bottom order.

Conditional independence relations in Bayesian networks...

- 2. A node is conditionally independent of all other nodes in the network, given its parents, children, and children's parents-that is, given its Markov blanket.
    - For example, *Burglary* is independent of *JohnCalls* and *MaryCalls*, given *Alarm* and *Earthquake*.



BAYESIAN NETWORK

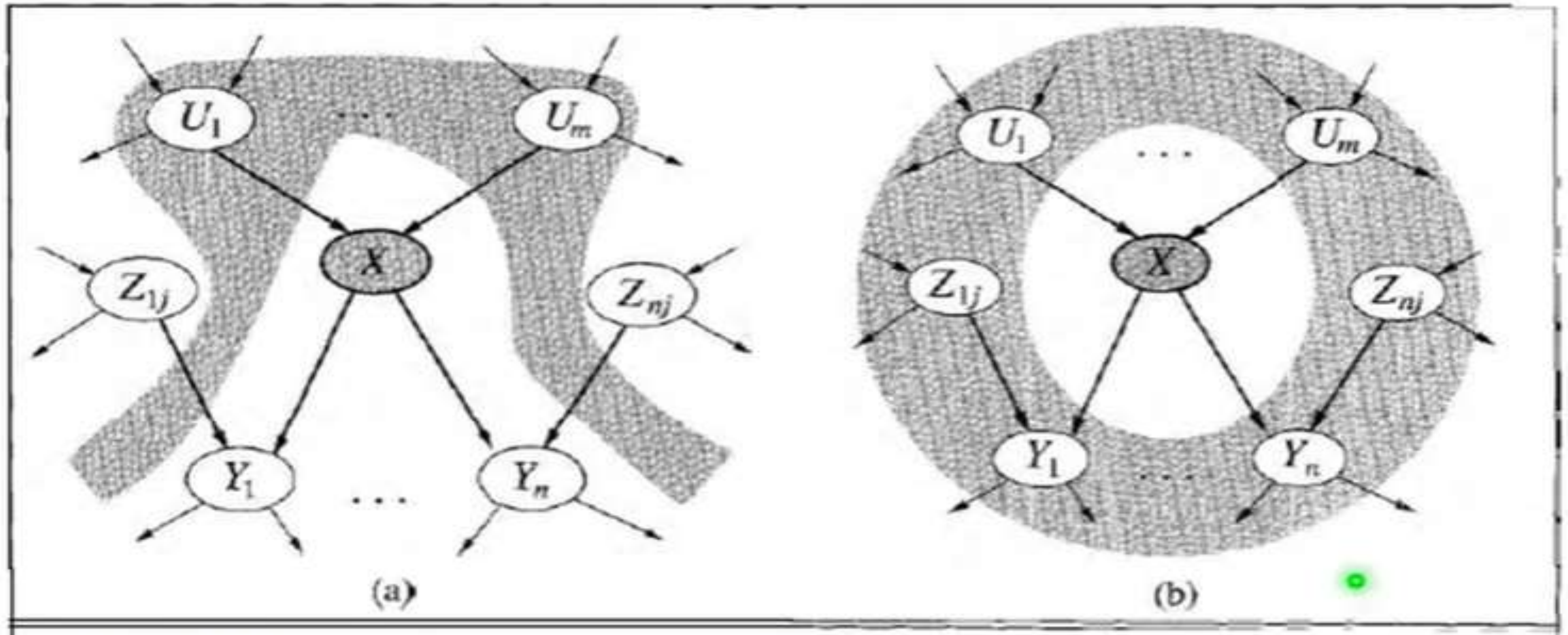


Figure 14.4 (a) A node X is conditionally independent of its non-descendants (e.g., the Z_{ij}s) given its parents (the U_is shown in the gray area). (b) A node X is conditionally independent of all other nodes in the network given its Markov blanket (the gray area).

14.3 Efficient Representation of Conditional Distributions

- CPT cannot handle large number or continuous value varibles.
- Relationships between parents and children are usually describable by some proper canonical distribution.
- Use the deterministic nodes to demonstrate relationship.
  - Values are specified by some function.
  - nondeterminism(no uncertainty)
  - Ex. X = f(parents(X))
  - Can be logical
    - NorthAmerica ↔ Canada v US v Mexico
  - Or numercial
    - Water level = inflow + precipitation – outflow - evaporation

14.3 Efficient Representation of Conditional Distributions

- Noisy-OR
  - All possible causes are listed. (the missing can be covered by leak node)
  - Compute probability from the inhibition probability

$$P(x_i \mid parents(X_i)) = 1 - \prod_{\{j: X_j = true\}} q_j \,,$$

14.3 Efficient Representation of Conditional Distributions

- Suppose these individual inhibition probabilities are as follows:

$$q_{\text{cold}} = P(\neg fever \mid cold, \neg flu, \neg malaria) = 0.6\,,$$
$$q_{\text{flu}} = P(\neg fever \mid \neg cold, flu, \neg malaria) = 0.2\,,$$
$$q_{\text{malaria}} = P(\neg fever \mid \neg cold, \neg flu, malaria) = 0.1$$

Cold	*Flu*	*Malaria*	*P(Fever)*	*P(¬Fever)*
F	F	F	0.0	1.0
F	F	T	0.9	**0.1**
F	T	F	0.8	**0.2**
F	T	T	0.98	$0.02 = 0.2 \times 0.1$
T	F	F	0.4	**0.6**
T	F	T	0.94	$0.06 = 0.6 \times 0.1$
T	T	F	0.88	$0.12 = 0.6 \times 0.2$
T	T	T	0.988	$0.012 = 0.6 \times 0.2 \times 0.1$

Variable depends on k parents can be described using O(k) parameters instead of O(2k)

Bayesian nets with continuous variables

- Many real world problems involve continuous quantities
  - Infinite number of possible values
  - Impossible to specify conditional probabilities
- Discretization
  - dividing up the possible values into a fixed set of intervals
  - It's often results in a considerable loss of accuracy and very large CPTs

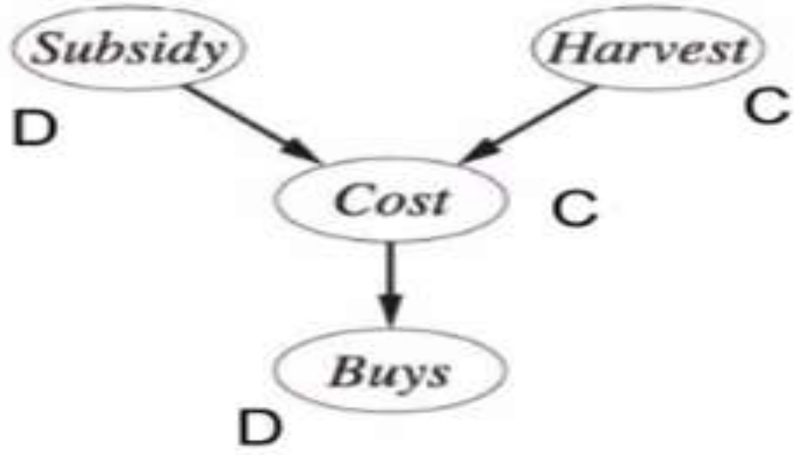
To define standard families of probability density functions(Gaussian,etc)

Bayesian nets with continuous variables

- Hybrid Bayesian network
  - Have both discrete and continuous variables
  - Two new kinds of distributions
    - Continuous variable given discrete or continuous parents
    - Discrete variable given continuous parents

- Example

Customer buys some fruit depending on its cost which depends in turn on the size of the harvest and whether the government's subsidy scheme is operating.



Hybrid Bayesian network

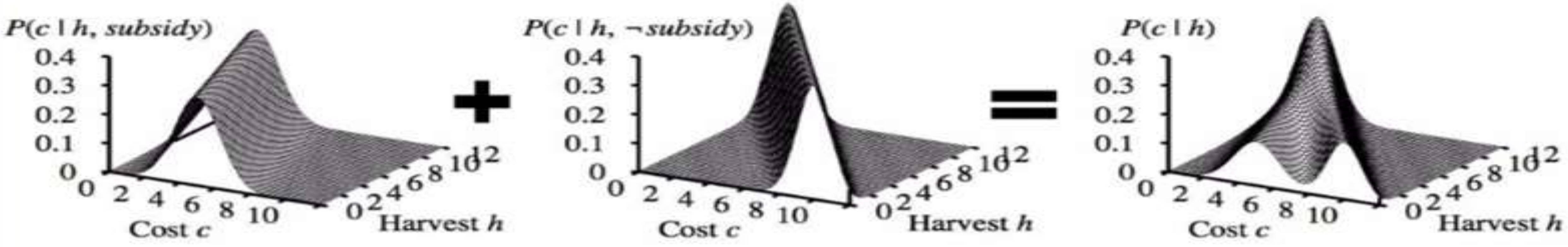
- The linear Gaussian distribution
  - Most common choice
  - The child has a Gaussian distribution
    - GD has μ varies linearly with the value of the parent
    - GD has standard deviation σ that is fixed
  - Two distributions, subsidy and ¬subsidy, with different parameters at, bt, σt, af , bf , and σf:

$$P(c \mid h, subsidy) = N(a_t h + b_t, \sigma_t^2)(c) = \frac{1}{\sigma_t\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{c-(a_t h+b_t)}{\sigma_t}\right)^2}$$

$$P(c \mid h, \neg subsidy) = N(a_f h + b_f, \sigma_f^2)(c) = \frac{1}{\sigma_f\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{c-(a_f h+b_f)}{\sigma_f}\right)^2}$$

Hybrid Bayesian network

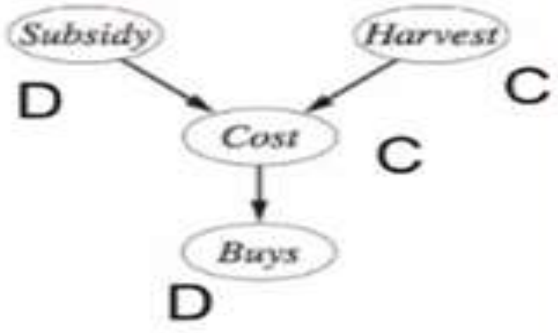
- A : P(Cost|Harvest,subsidy)
- B : P(Cost|Harvest,¬subsidy)
- C : P (c | h)
  - averaging over the two possible values of Subsidy
  - assuming that each has prior probability 0.5



(a) (b) (c)

Other Distribution

- Distributions for discrete variables with continuous parents
- Consider the Buys node
  - Customer will buy if the cost is low
  - Customer will not buy if it is high
  - the probability of buying varies smoothly

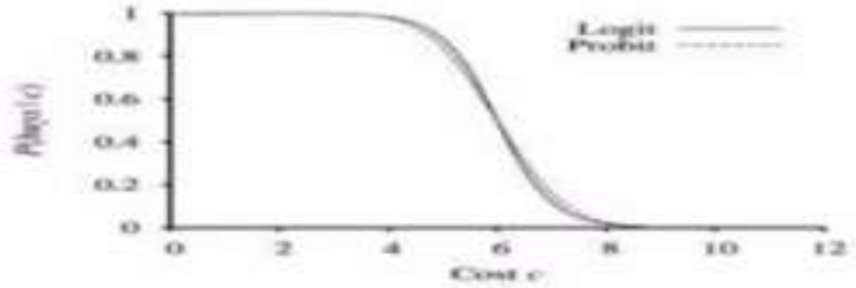


- Probit Distribution

$$\Phi(x) = \int_{-\infty}^{x} N(0,1)(x)dx \qquad P(buys \mid Cost = c) = \Phi((-c+\mu)/\sigma)$$

- Logit Distribution

$$P(buys \mid Cost = c) = \frac{1}{1 + exp(-2\frac{-c+\mu}{\sigma})}$$



THANK YOU