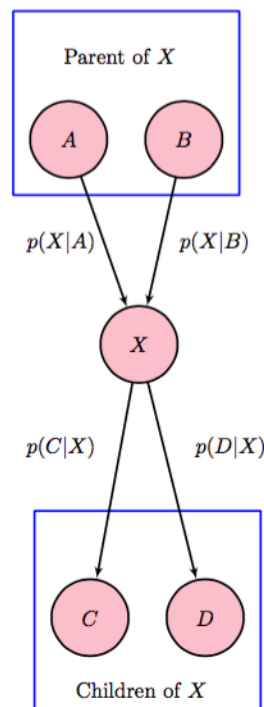


# Lecture 11

## Bayesian Networks

A Bayesian network is a directed acyclic graph that represents a probabilistic graphical models. In particular a Bayesian network is defined as a pair  $B = (G, \Theta)$  where  $G$  is a directed graph with no cycles whose vertices  $X_1, \dots, X_n$  represents random variables, whose edges represent the direct dependencies between these variables. The symbol  $\Theta$  represents the set of parameters specifies the probability distributions associated with each variable.

An edge from vertex  $X_i$  to  $X_j$  represents a statistical dependence between the corresponding variables. Vertex  $X_i$  is referred to as a parent of  $X_j$  and  $X_j$  is called the child of  $X_i$ . An extension of these genealogical terms is often used to define the sets of descendants which is the set of all vertices that can be reached on a direct path from the vertex. Ancestor vertices is vertices from which the vertex can be reached on a direct path.



## 11.1 Conditional Independence

The two variables  $X$  and  $Y$  are conditionally independent given  $Z$  if and only if

$$\mathbf{P}(X, Y|Z) = \mathbf{P}(X|Z)\mathbf{P}(Y|Z)$$

which is equivalent to

$$\mathbf{P}(X|Y, Z) = \mathbf{P}(X|Z) \text{ or } \mathbf{P}(Y|X, Z) = \mathbf{P}(Y|Z)$$

Recall the joint probability of  $X_1, \dots, X_n$ :

$$p(X_1, \dots, X_n) = \prod_{i=1}^n p(X_i|X_1, \dots, X_{i-1})$$

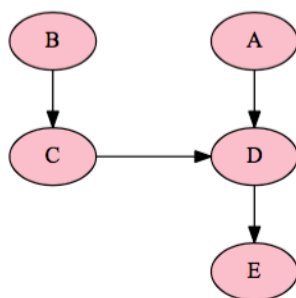
using the probability chain rule.

In Bayesian network, the state  $X_i$  is conditionally independent of its ancestors given its parents. Therefore,

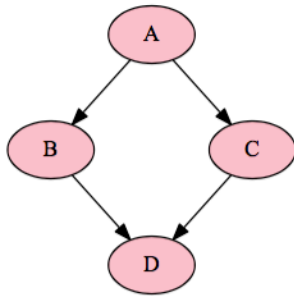
$$\mathbf{P}(X_1, \dots, X_n) = \prod_{i=1}^n \mathbf{P}(X_i|\text{parents of } X_i)$$

**Exercise 11.1** Write the joint probability distribution of the following Bayesian networks.

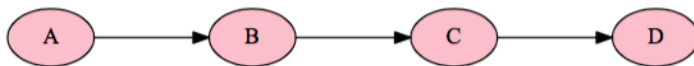
1.



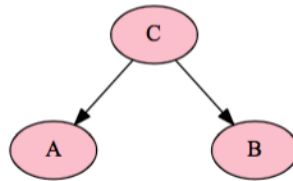
2.



3.



## 11.2 Common causes (tail to tail)



The vertex  $C$  is said to be tail-to-tail w.r.t. the path from vertex  $A$  to vertex  $B$ .

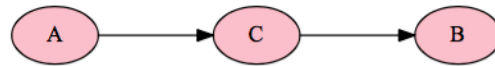
$$p(A, B, C) = p(A|C)p(B|C)p(C)$$

$$p(A, B) = \sum_{c \in C} p(A|c)p(B|c)p(c) \neq p(A)p(B)$$

$$p(A, B|C) = \frac{p(A|C)p(B|C)p(C)}{p(C)} = p(A|C)p(B|C)$$

$A$  and  $B$  are not independent but they are conditionally independent, given  $C$ .

### 11.3 Causal chain (head to tail)



The vertex  $C$  is said to be head-to-tail w.r.t. the path from vertex  $A$  to vertex  $B$ .

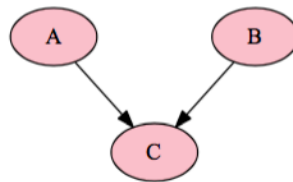
$$p(A, B, C) = p(A)p(C|A)p(B|C)$$

$$p(A, B) = \sum_{c \in C} p(A)p(c|A)p(B|c) \neq p(A)p(B)$$

$$p(A, B|C) = \frac{p(A)p(C|A)p(B|C)}{p(C)} = p(A|C)p(B|C)$$

$A$  and  $B$  are not independent but they are conditionally independent, given  $C$ .

### 11.4 Common effects (head to head)



The vertex  $C$  is said to be head-to-head w.r.t. the path from vertex  $A$  to vertex  $B$ .

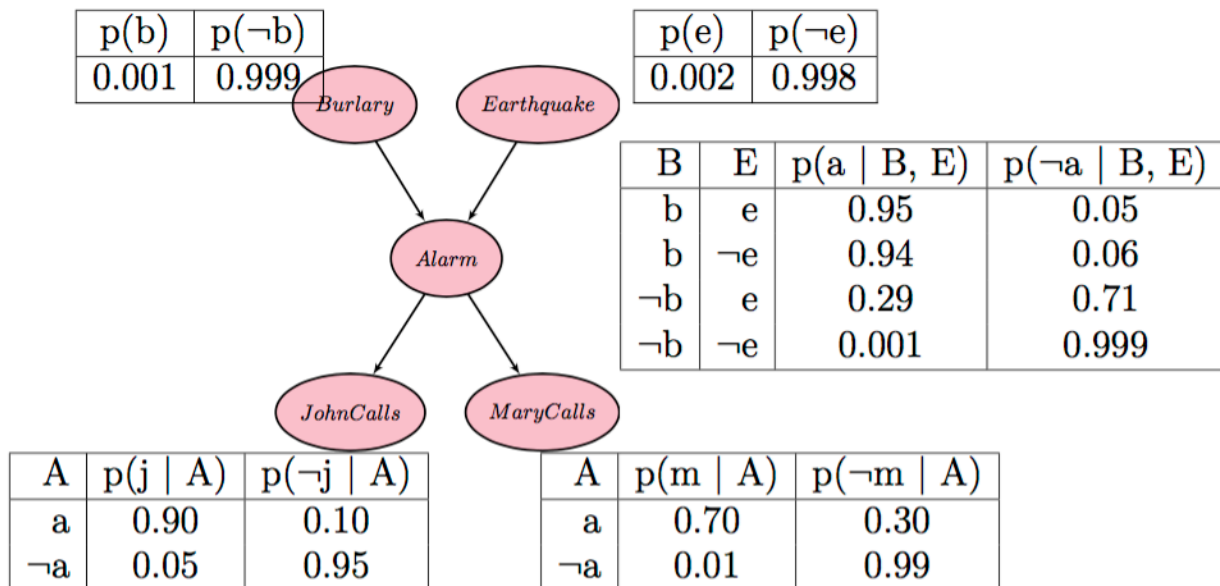
$$p(A, B, C) = p(A)p(B)p(C|A, B)$$

$$p(A, B) = \sum_{c \in C} p(A)p(B)p(c|A, B) = p(A)p(B) \sum_{c \in C} p(c|A, B) = p(A)p(B)$$

$$p(A, B|C) = \frac{p(A)p(B)p(C|A, B)}{p(C)} \neq p(A|C)p(B|C)$$

$A, B$  are independent, but not conditionally independent given  $C$ .

**Exercise 11.2** You have a new burglar alarm installed. It reliably detects burglary, but also responds to minor earthquakes. Two neighbors, John and Mary, promise to call the police when they hear the alarm. John always calls when he hears the alarm, but sometimes confuses the alarm with the phone ringing and calls then also. On the other hand, Mary likes loud music and sometimes doesn't hear the alarm. Given evidence about who has and hasn't called, you'd like to estimate the probability of a burglary. A BN representation is shown below (Pearl 1988)



1. What is the probability that the alarm has sounded but neither a burglary nor an earthquake has occurred, and both Mary and John call?

2. What is the probability that the alarm would sound, given that there was an earthquake?

3. What is the probability that the alarm will sound?

4. What is the probability that there was a burglary, given that John called?



**Exercise 11.3** “A Lecturer’s Life.” Dr. Ann Nicholson spends 60% of her work time in her office. The rest of her work time is spent elsewhere. When Ann is in her office, half the time her light is off (when she is trying to hide from students and get research done). When she is not in her office, she leaves her light on only 5% of the time. 80% of the time she is in her office, Ann is logged onto the computer. Because she sometimes logs onto the computer from home, 10% of the time she is not in her office, she is still logged onto the computer.

1. Construct a Bayesian network to represent the “Lecturer’s Life” scenario just described.

2. Suppose a student checks Dr. Nicholson's login status and sees that she is logged on. What effect does this have on the student's belief that Dr. Nicholson's light is on?

## Credit

This lecture note is partially adopted from Asst. Prof. Dr. Nirattaya Khamsemanan's GTS111 and ICT600 partial notes.