

[Home](#) [Read](#) [Sign](#)in [Search in book ...](#)

---

## MACHINE LEARNING

### CONTENTS

# 30. Bayesian Belief Networks-II

### Learning Objectives:

The learning objectives of this module are as follows:

- To understand the design of the Bayesian network
- To understand the examples of Bayesian networks
- To explain inferencing in Bayesian Belief Networks

### 32.1 Constructing Bayesian networks

Let us first discuss the construction of a Bayesian network. Let us assume that the problem can be defined in terms of  $n$  random variables. The steps in the construction is given below:

1. Choose an ordering of variables  $X_1, \dots, X_n$

Increase Font Size

2. For  $i = 1$  to  $n$ add  $X_i$  to the networkselect parents from  $X_1, \dots, X_{i-1}$  such that

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

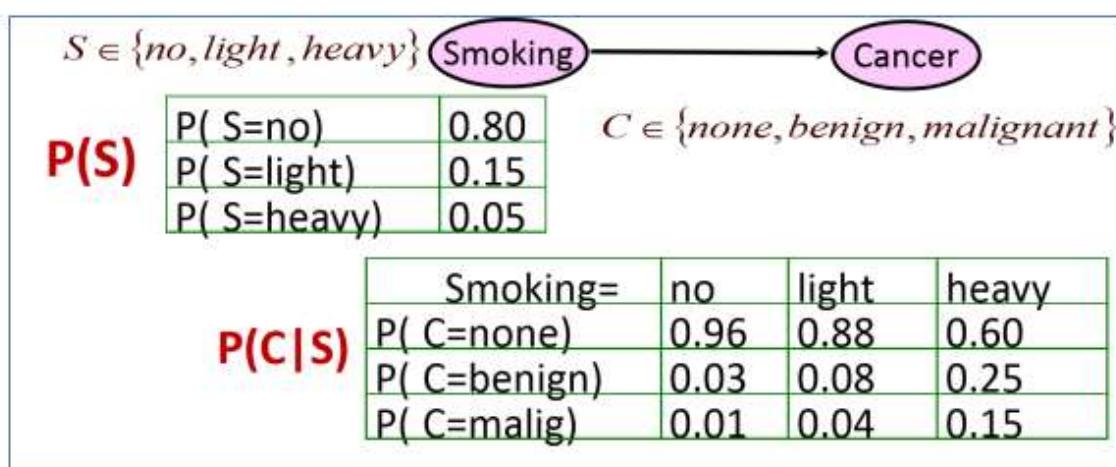
This choice of parents guarantees:

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

$$= \prod_{i=1}^n P(X_i | \text{Parents}(X_i)) \text{ (by construction)}$$

### 32.2 Bayesian Networks – Smoking – Example I

Let us consider the simple example given below:



**Figure 32.1 Example of Bayesian Network**

The example shows Cancer having Smoking as parent where Smoking and Cancer each can take three values. Probability of Smoking being each of the three values that is Probability of smoking being no, Probability of smoking being light, and Probability of smoking being heavy are given. The Conditional Probability Table (CPT) of each of the three values of Cancer (namely none, benign, malignant) given the different values of Smoking is given (Figure 32.1). The product rule for the example is shown in Figure 32.2. Here the conditional probability  $P(C|S)$  is multiplied by

probability of  $s$   $P(S)$  to obtain joint probability  $P(C,S)$ . This table shows marginalization as shown in Figure 32.3. The total of the columns gives us  $P(C)$  – the marginalized value of  $C$ .

[Increase Font Size](#)

$P(C,S) = P(C S) \cdot P(S)$				
$S \downarrow$	$C \Rightarrow$	none	benign	malignant
no	0.768	0.024	0.008	
light	0.132	0.012	0.006	
heavy	0.035	0.010	0.005	

Figure 32.2 Product Rule for the Example

$S \downarrow$	$C \Rightarrow$	none	benign	malignant	total	
no	0.768	0.024	0.008	0.800		P(Smoke)
light	0.132	0.012	0.006	0.150		
heavy	0.035	0.010	0.005	0.050		
total	0.935	0.046	0.019			

Figure 32.3 Marginalization

Now using the values calculated we get  $P(S|C) = P(C|S) \cdot P(S) / P(C) = P(C,S) / P(C)$  as shown in Figure 32.4. Therefore we get the probability of the state of Smoking knowing the state of Cancer can be inferred.

$P(S C) = P(C S) \cdot P(S) / P(C) = P(C,S) / P(C)$				
$S \downarrow$	$C \Rightarrow$	none	benign	malignant
no	0.768/.935	0.024/.046	0.008/.019	
light	0.132/.935	0.012/.046	0.006/.019	
heavy	0.035/.935	0.010/.046	0.005/.019	
Cancer=		none	benign	malignant
P(S=no)	0.821	0.522	0.421	
P(S=light)	0.141	0.261	0.316	
P(S=heavy)	0.037	0.217	0.263	

Figure 32.4 Inferred Values using Marginalization

Increase Font Size

### 32.3 Causes and Bayes' Rule

Let us consider a simple example of Rain causing Wet-Grass (Figure 32.5). We are given that the probability of Rain  $P(R)$  is 0.4 which also means that the probability of no rain is 0.6. In addition we are given the CPT of the Wet-Grass under the conditions of Rain and No Rain. Now how do we proceed to carry out diagnostic inference that is Knowing that the grass is wet, what is the probability that rain is the cause?

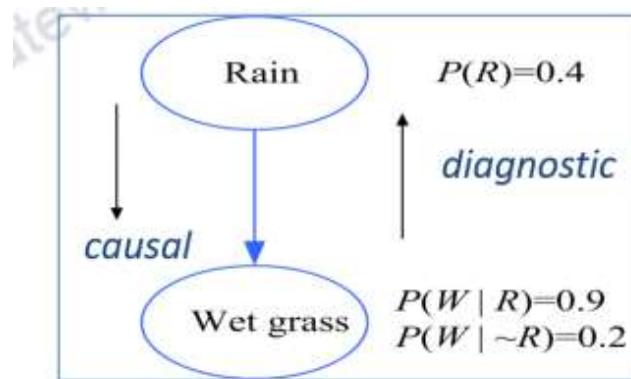


Figure 32.5 Causal and Diagnostic Inference

The procedure for carrying out this diagnostic inference is shown in Figure 32.6. Now using Bayes Theorem we know that  $P(R|W)$  can be deduced from  $P(W|R)$

\*  $P(R)$  divided by  $P(W)$ . Now probability of Wet-Grass  $P(W)$  is the summation of  $P(W|R)*P(R) + P(W|~R)*P(~R)$  that is the probability of Wet-Grass is based on the probabilities of the states of its cause (Rain), in this case only two states R and  $\sim R$ .

$$\begin{aligned}
 P(R | W) &= \frac{P(W | R)P(R)}{P(W)} \\
 &= \frac{P(W | R)P(R)}{P(W | R)P(R) + P(W | \sim R)P(\sim R)} \\
 &= \frac{0.9 \times 0.4}{0.9 \times 0.4 + 0.2 \times 0.6} = 0.75
 \end{aligned}$$

**Figure 32.6 Diagnostic Inference****Increase Font Size**

### 32.4 Bayesian Networks Construction – Icy Roads – Example II

Let us now understand how to design a Bayesian Network for the problem statement given in Figure 32.7 .

1. Inspector Smith is waiting for Holmes and Watson who are both late for an appointment.
2. Smith is worried that if the roads are icy one or both of them may have crashed his car.
3. Suddenly Smith learns that Watson has crashed.
4. Smith Thinks – If Watson has crashed, probably roads are icy, then Holmes has probably crashed too.
5. Smith then learns it is warm outside and roads are salted.
6. Smith thinks – Watson was unlucky, Holmes should still make it.

**Figure 32.7 Icy Roads – Problem Statement**

We can model this problem using three random variables namely State of Roads Icy/not Icy, Watson Crash/no Crash and Holmes Crash/no Crash. The Bayesian Network has these three nodes with State of Roads being the reason for Watson Crash/no Crash and Holmes Crash/no Crash (Figure 32.8). Now let us understand the flow of information in the network. Now in sentence 3 in the problem statement (Figure 32.7) Smith knows that Watson has crashed. Therefore Watson Crash is true and now information flows from this node to the State of Roads node. Now in sentence 5 Smith comes to know that it is warm outside and roads are salted. Therefore now information flows from State of Roads node to Holmes Crash/no Crash node. This example shows how different types of inferences needed depends on the knowledge acquired.

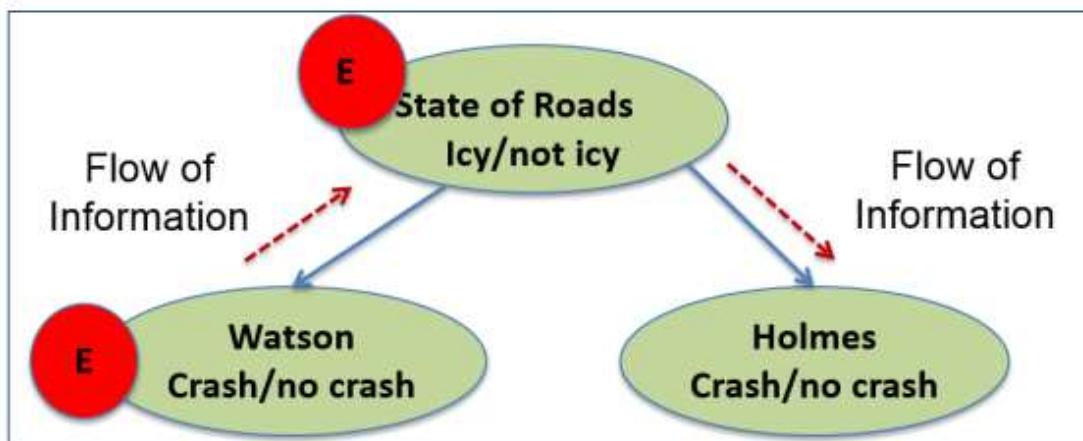


Figure 32.8 Bayesian Network for Icy Roads Exam

Increase Font Size

### 32.5 Conditional Independence in BNs

Normally edges between the nodes of a Bayesian Network represent conditional dependencies. Nodes that are not connected that is where there is no path from a node to another in the Bayesian network represent variables that are said to be conditionally independent of each other. There are three types of conditional independences associated with Bayesian Networks, namely serial, diverging and converging (Figure 32.9).

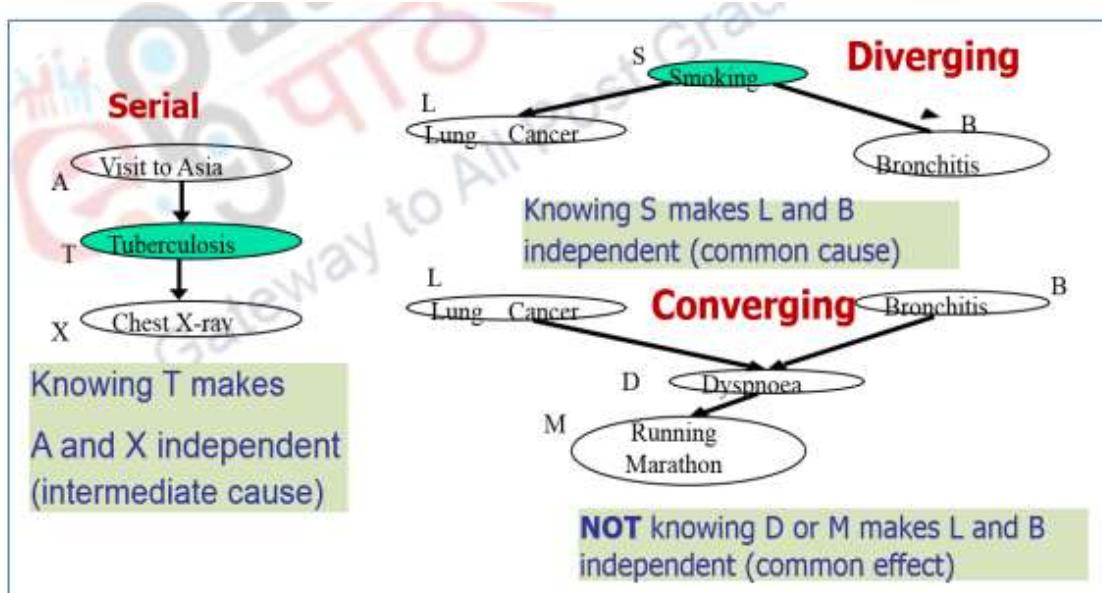


Figure 32.9 Conditional Independence

#### 32.5.1 Serial Case

In the case of serial we have T and given the condition that T is known, A and X are conditionally independent. Conditional independence is due to the fact that the intermediate cause is known.

#### 32.5.2 Diverging Case

In the case of diverging we have S being the common cause for two nodes L and B that is the two nodes L and B is connected through the common cause S. Now given

the condition that S is known, L and B are conditionally independent. Independence is due to the fact that the common cause is known.

### 32.5.3 Converging Case

In the case of diverging we have two nodes L and B both being causes for the node D which in turn is the cause for the node M. Now given the condition that D is known, the common effect of L and B is not known, and neither is the effect M of the common effect known, then L and B are conditionally independent. Conditional independence is due to the fact that the common effect and in turn its effect is not known.

## 32.6 Bayesian Networks – Alarm (from Judea Pearl) – Example III

The alarm example is a good example to explain many aspects of Bayesian Networks and is therefore a very popular example. Here we use the example to explain the steps in the construction of a Bayesian Network. The problem statement is given in Figure 32.10. The steps in the construction are given below:

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

### Figure 32.10 Bayesian Network – Alarm Example

#### 32.6.1 Step 1:

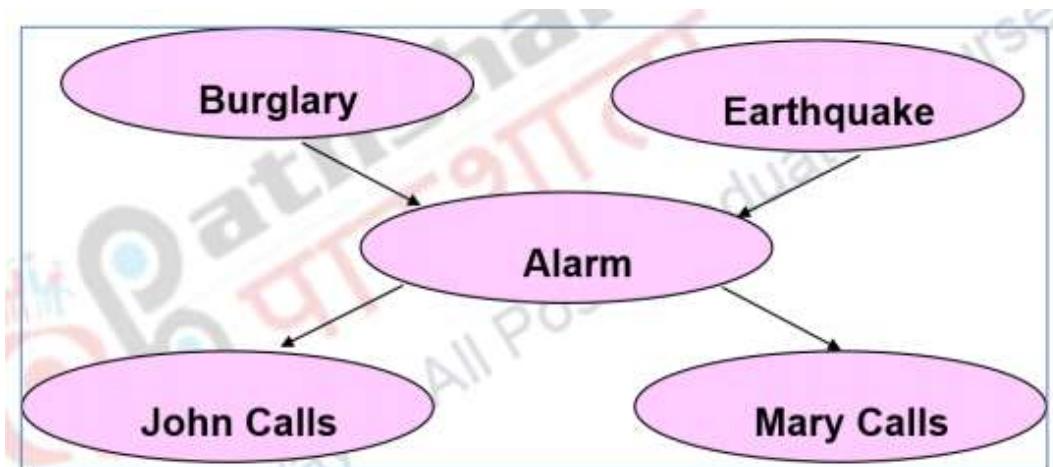
First we determine what the propositional (random) variables should be. Then we determine causal (or another type of influence) relationships and develop the topology of the network.

Increase Font Size

Variables are identified as: *Burglary*, *Earthquake*, *Alarm*, *JohnCalls*, *MaryCalls* The Network topology reflecting the “causal” knowledge is as follows:

- A burglar can set the alarm off
- An earthquake can set the alarm off
- The alarm can cause Mary to call
- The alarm can cause John to call

The resulting Topology of the Bayesian Network is shown in Figure 32.11.



**Figure 32.11 Topology of the Network**

### 32.6.2 Step 2:

Next we need to specify a *conditional probability table* or CPT for each node. This can be done through observations or using heuristics. Each row in the table contains the conditional probability of each node value for a conditioning case (possible combinations of values for parent nodes). In the example, the possible values for each node are true/false. The sum of the probabilities for each value of a node given a particular conditioning case is 1. For example the CPT for *Alarm* Node is shown in Figure 32.12. This CPT indicates the probability of *Alarm* given *Burglar* and *Earthquake*. Now both *Burglar* and *Earthquake* can each take values of T/F. Hence

Increase Font Size

this CPT has four rows. The value of both T and F values of  $P(A|B)$  and  $P(A|E)$  values of B and E is indicated in the CPT (Figure 32.12).

B	E	$P(A B,E)$	
		True	False
T	T	0.950	0.050
T	F	0.940	0.060
F	T	0.290	0.710
F	F	0.001	0.999

Figure 32.12 CPT of Alarm Node

Similarly the CPT for each of the other nodes is shown in Figure 32.13. Please note that the nodes Burglar and Earthquake are independent and hence independent probability is associated with these nodes.

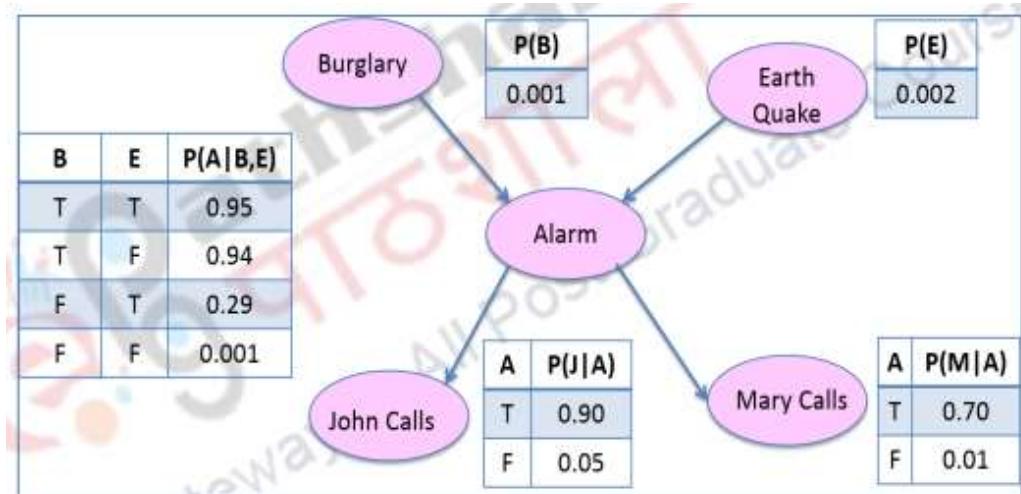


Figure 32.13 Bayesian Network with CPTs of all Nodes for Alarm Example

## 32.7 Semantics of Bayesian Networks

There are two views of a Bayesian Network. View 1 says that a belief network is a representation of the joint probability distribution (“joint”) of a domain. The joint

**Increase Font Size**

completely specifies an agent's probability assignments to all domain (both simple and complex). In other words the network can be viewed as a representation of joint probability distribution of its variables. A generic entry in the joint probability distribution is the probability of a conjunction of particular assignments to each variable, such as:

$$P(x_1, \dots, x_n) = \prod_{i=1}^n P(x_i | \text{Parents}(X_i))$$

We can see that each entry in the joint is represented by the product of appropriate elements of the CPTs in the belief network.

For the Alarm example we can calculate the probability of the event that the alarm has sounded but neither a burglary nor an earthquake has occurred, and both John and Mary call as given below:

$$\begin{aligned} & P(J \wedge M \wedge A \wedge \neg B \wedge \neg E) \\ &= P(J|A) P(M|A) P(A|\neg B, \neg E) P(\neg B) P(\neg E) \\ &= 0.90 * 0.70 * 0.01 * 0.999 * 0.998 \\ &= 0.00062 \end{aligned}$$

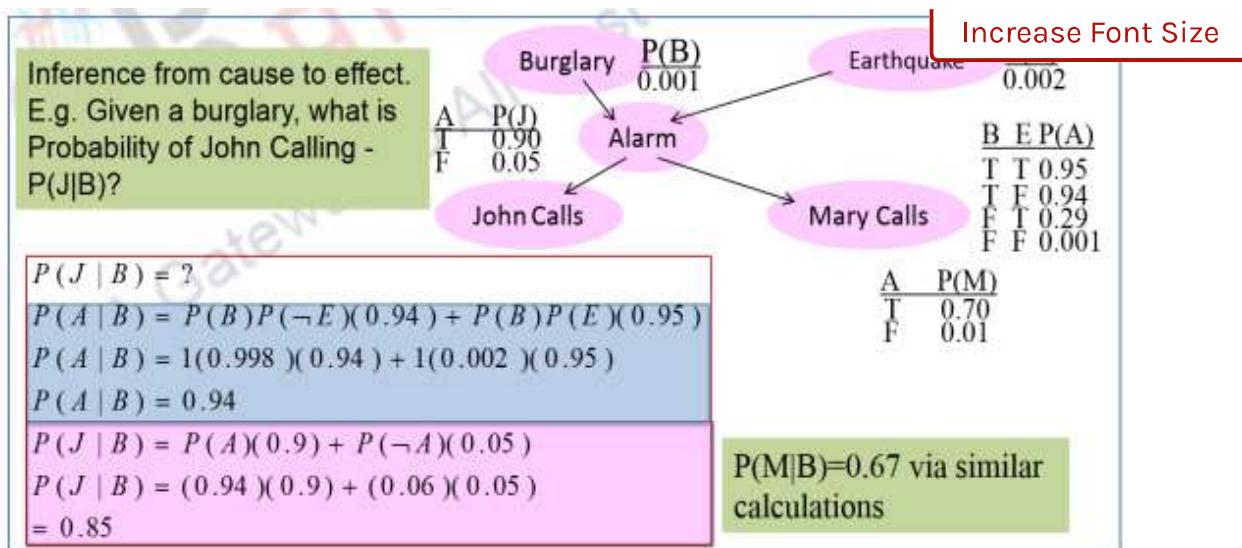
The second view is that the Bayesian network is an encoding of a collection of conditional independence statements. For the Alarm example

- **JohnCalls is conditionally independent of other variables in the network given the value of Alarm**

This view is useful for understanding inference procedures for the networks .

### 32.8 Inferences in Bayesian Networks

Basically as we have already discussed there are two basic types of inferences associated with Bayesian Networks namely causal Inferences and diagnostic inferences.



**Figure 32.14 Causal Inference**

### 32.8.1 Causal Inference:

Causal Inference is basically inference from cause to effect. In Figure 32.14 given Burglary we can find the probability of John calling ( $P(J|B)$ ). Note that this is indirect inference since Burglary causes Alarm which in turn causes John to call. Therefore we first need the probability of Alarm ringing given that Burglary has occurred ( $P(A|B)$ ).

#### 32.8.1.1 Step 1: Calculating $P(A|B)$

$P(A|B)$  is calculated by considering the probability of Alarm ringing based on both occurrence of Burglary (probability 1 since this is given) and occurrence of Earthquake (which may or may not occur). Therefore for  $P(A|B)$  we consider the two rows from the CPT of Alarm node where Burglary is True and calculate  $P(A|B)$  as the sum of these two terms with associated probabilities as given in Figure 32.14.

#### 32.8.1.2 Step 2: Calculating $P(J|B)$

In the next step we need to calculate probability of John calling given the probability of Burglary. Here we have that either the Alarm rings when burglary occurs  $P(A)$  which we calculated in Step 1 or does not ring which is  $1 - P(A)$ . Now we associate the probability of  $P(J)$  given  $A$  from the CPT of  $J$  as given in Figure 32.13. Using these two probability values we can calculate the probability of John calling given burglary occurred (Figure 32.14).

## 32.8.2 Diagnostic Inference

Diagnostic Inference is basically inference from effect to cause. In Figure 32.15 given that John calls, we can find the probability of Burglary occurring ( $P(B|J)$ ). Note that this is indirect inference since John calling was caused by Alarm ringing  $P(J|A)$  which in turn was caused by a Burglary ( $P(B|A)$ ).

Now we first apply Bayes theorem to find  $P(B|J)$  as given in Figure 32.15. Now this shows that we need to know the probability of John calling that is  $P(J)$ . However for finding probability of  $P(J)$  we need to know probability of Alarm ringing that is  $P(A)$ .

### 32.8.2.1 Step 1: Calculating $P(A)$

For calculating  $P(A)$ , we do not have any information of whether Burglary occurred or Earthquake occurred. Hence we consider the four rows of the CPT of Alarm node with probability of truth values of Burglary and Earthquake from their respective probability tables (Figure 32.15).

### 32.8.2.2 Step 2: Calculating $P(J)$

We use the value of  $P(A)$  calculated in step 1 for the calculation of probability of John calling that is  $P(J)$ . Here we consider the probability that Alarm rang and did not ring ( $1 - P(A)$ ) for the calculation of  $P(J)$  from its CPT.

### 32.8.2.3 Step 3: Calculating $P(B|J)$

Now we can calculate the probability of Burglary given John called using the Bayes theorem. We first need the value of  $P(J|B)$  which we can calculate as shown in Section 32.8.1. The probability value of occurrence of Burglary  $P(B)$  can be obtained from its probability table and  $P(J)$  has been calculated as explained in Section 32.8.2.2.

$$P(B | J) = \frac{P(J | B)P(B)}{P(J)}$$

What is P(J)? Need P(A) first.

$$P(A) = P(B)P(E)(0.95) + P(\neg B)P(E)(0.29)$$

$$+ P(B)P(\neg E)(0.94) + P(\neg B)P(\neg E)(0.001)$$

$$P(A) = (0.001)(0.002)(0.95) + (0.999)(0.002)(0.29)$$

$$+ (0.001)(0.998)(0.94) + (0.998)(0.999)(0.001)$$

$$P(A) = 0.002517$$

$$P(J) = P(A)(0.9) + P(\neg A)(0.05)$$

$$P(J) = (0.002517)(0.9) + (0.9975)(0.05)$$

$$P(J) = 0.052$$

$$P(B | J) = \frac{(0.85)(0.001)}{(0.052)} = 0.016$$

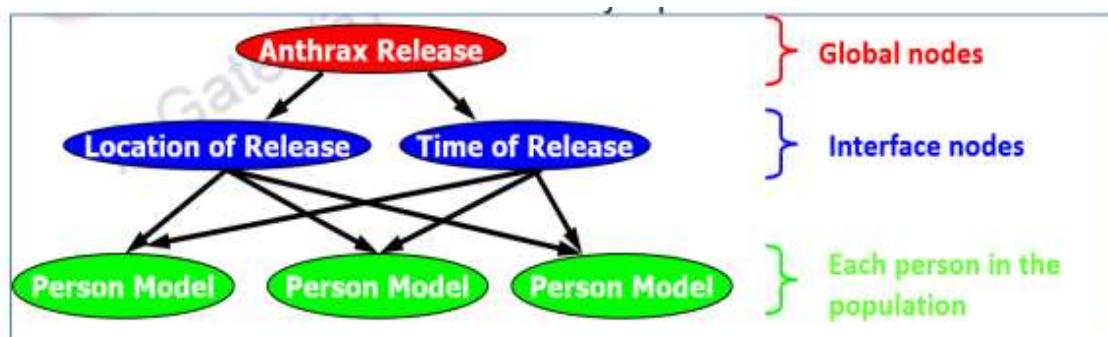
Increase Font Size

**Figure 32.15 Diagnostic Inference**

### 32.9 Realistic Example – Population-wide ANomaly Detection & Assessment (PANDA)

A realistic example of the use of Bayesian network is a detector specifically for a large-scale outdoor release of inhalational anthrax (reference Bayesian Biosurveillance of Disease Outbreaks, Gregory F., Denver H. Dash, John D.

Levander, Weng-Keen Wong, Proceedings of the 20th conference on Uncertainty in Artificial Intelligence, 2004). Here a massive causal Bayesian network was used with a population-wide approach where each person in the population is represented as a sub-network in the overall model (Figure 32.16). Please note the conditional independence assumptions. Also note that Anthrax is assumed to be infectious but non-contagious. The structure of the Network that is the causal dependencies were designed through expert judgment and the parameters obtained from census data, training data, and expert assessments obtained from literature and by experience.



**Figure 32.16 Panda Network****Increase Font Size****Summary**

- Explained the construction of the Bayesian network
- Discussed the examples of Bayesian networks
- Outlined inferencing in Bayesian Belief Networks.

**Web Links**

- <http://people.cs.pitt.edu/~milos/courses/cs2750-Spring03/lectures/class2.pdf>
- <http://learningforward.org/docs/default-source/commoncore/comprehensive-professional-learning system.pdf?type=application/pdf>
- [http://digitalcommons.ilr.cornell.edu/cgi/viewcontent.cgi?article=1405&context=c\\_ahrswp](http://digitalcommons.ilr.cornell.edu/cgi/viewcontent.cgi?article=1405&context=c_ahrswp) type="application/pdf"
- <http://www.cse.hcmut.edu.vn/~tru/AI/chapter11.pdf> type="application/pdf"
- [http://www.holehouse.org/mlclass/11\\_Machine\\_Learning\\_System\\_Design.html](http://www.holehouse.org/mlclass/11_Machine_Learning_System_Design.html)
- <ssdi.di.fct.unl.pt/pc/0607/files/PCaulaT03 -10-06.ppt>

**Supporting & Reference Materials**

- Tom [Mitchell](#), “Machine Learning”, McGraw-Hill Education, 1997
- [AlpaydinEthem](#), “Introduction to Machine Learning”, The MIT Press; third edition, 2014
- Christopher M. [Bishop](#), “Pattern Recognition and Machine Learning”, Springer, 2013 [Peter Harrington](#), “Machine Learning In Action”, Manning Publications, 2012
- [Peter Flach](#), “Machine Learning: The Art and Science of Algorithms that Make Sense of Data”, Cambridge University Press, 2012
- [Stephen Marsland](#), “Machine Learning: An Algorithmic Perspective”, Chapman and Hall/CRC; 2 edition, 2014
- Y. S. Abu-Mostafa, M. Magdon-Ismail, and H.-T. Lin, “Learning from Data”, AMLBook

Increase Font Size

Machine Learning Copyright © by Ranjani Parthasarathi. All Rights Reserved.

