Lecture 12
Bayesian Networks

Alice Gao

November 3, 2021

Contents

Learning Goals

By the end of the lecture, you should be able to

- Given a Bayesian network, determine whether an (conditional) independence relationship holds using d-separation.
- Given a joint probability distribution and an order of the variables, construct a Bayesian network that correctly represents the independent relationships among the variables in the distribution.

1 Testing independence using d-separation

Up to now, we haven't had the tools to test whether an independence relationship holds. Let me present a powerful concept called d-separation, which we can use to determine whether an unconditional or conditional independence relationship holds or not.

1.1 D-Separation Definition

See below for the definition of d-separation. There are two variables X and Y, and a set of observed variables E.

Definition (D-Separation)**.** A set of variables E **d-separates** variables X and Y if E blocks every un-directed path between X and Y in the network.

To determine whether X and Y are independent given the observed variables E, we can verify whether E d-separates X and Y. If d-separation holds, then the independence relationship holds as well.

Let me clarify a few things regarding the d-separation definition.

First, to verify the definition, we need to consider every un-directed path between X and Y. The word "un-directed" means that we do not care about the direction of the arrows on the path. As long as a series of nodes and edges connect X and Y, we will consider it a path.

Second, there may be multiple paths between X and Y. We need to consider every path and verify that E blocks every path between X and Y.

Third, on each path, there could be multiple nodes between X and Y. The path is blocked if at least one node blocks the path. As soon as we find one node blocking the path, we can move on to a different path. In the worst case, we need to check every node and discover that none of the nodes blocks the path.

Given this definition, our task boils down to the following: pick a path between X and Y and pick a node on the path, determine whether the node blocks the path or not.

This leads to our next question: What does it mean to "block a path"? Let me explain this in three scenarios. Interestingly, these three scenarios correspond to the three key structures that I discussed previously.

In each of the three scenarios, we will look at one path between X and Y and consider one node N on the path. The three scenarios differ by the direction of the two arrows on both sides of N.

1.2 Blocked Path — 3 Scenarios

Scenario one:

The two arrows around N point in the same direction, forming a chain around N. I drew the arrows pointing to the right, but it's fine if they point to the left. In this scenario, if N is observed, then N blocks the path between X and Y.

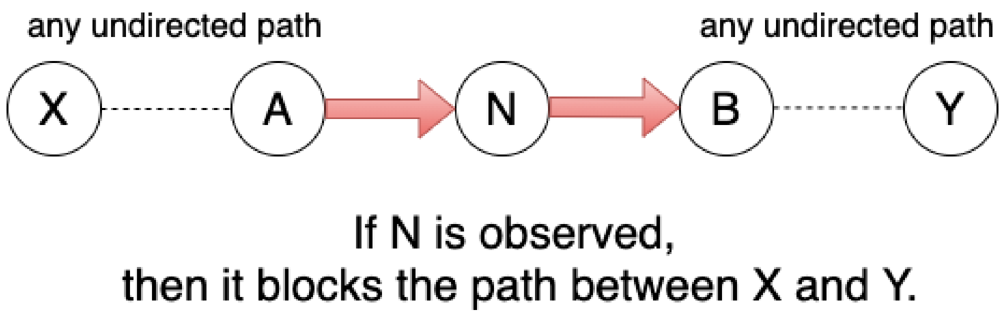


Figure 1: Scenario one

This rule is similar to the first key structure that is a chain. If we observe whether Alarm is going off, then Burglary and Watson become independent. You can think of observing N as cutting the chain at N.

Scenario two:

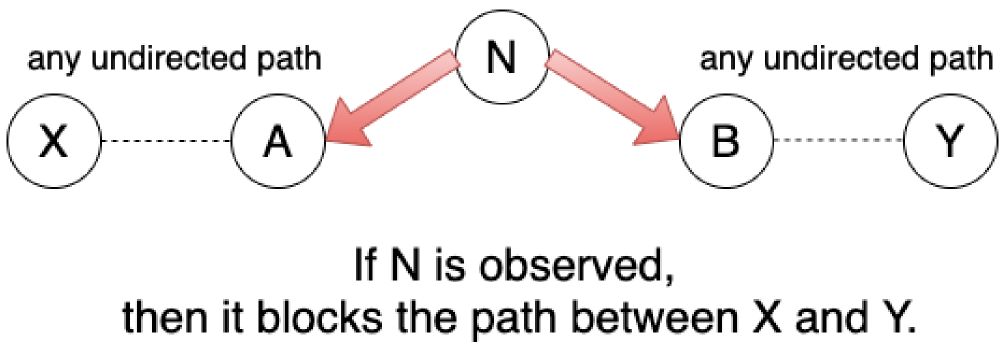


Figure 2: Scenario two

The two arrows around N point away from N, to the two children A and B. If the arrows depict causal relationships, you can think of A and B as unreliable sensors of N. If N is observed, then N blocks the path between X and Y.

This rule is similar to the second key structure. If we observe Alarm, then Watson and Gibbon become independent.

Scenario three:

The two arrows around N point toward N. A and B are both parents of N. If the arrows depict causal relationships, then A and B jointly cause N to happen. The descendants of N are also important in this scenario. The rule says that: If we do not observe N and do not observe any of N's descendants, then the path is blocked.

This rule is similar to the third key structure. If Alarm is not observed, then Burglary and Earthquake are independent. If Alarm is observed, Burglary and Earthquake become dependent.

Note that this rule is the opposite of the first two rules. The first two rules say that N blocks

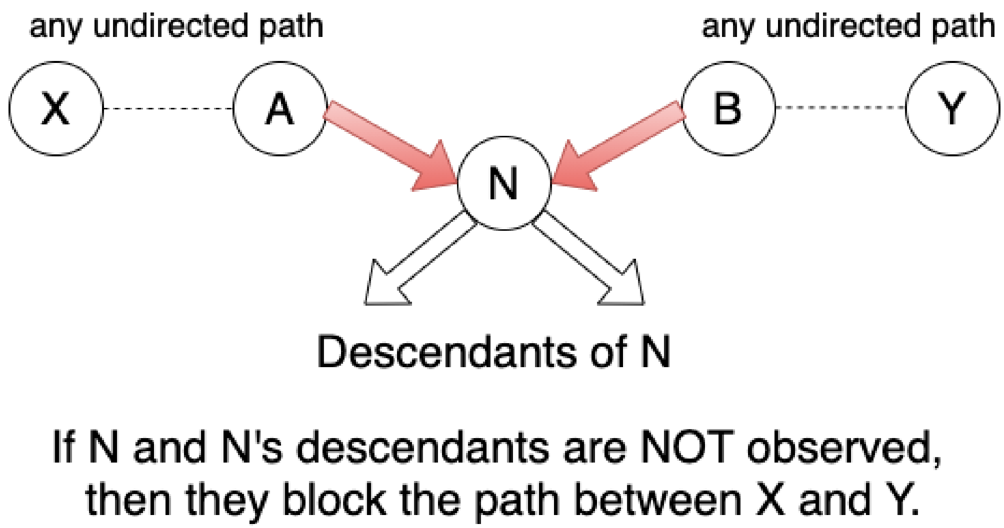


Figure 3: Scenario three

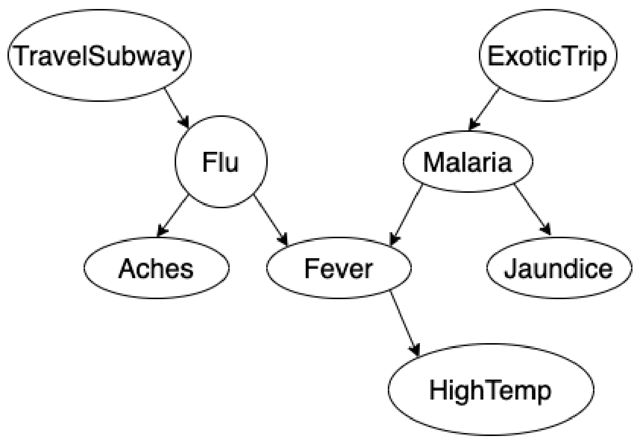
the path if N is observed. This third rule says that N and its descendants block the path if they are not observed.

1.3 Examples of Applying D-Separation

We will use the same Bayesian network for all the examples. The network describes the following scenario.

If I travel on the subway, I may get the flu. If I take a trip to an exotic destination, I may catch malaria. Flu and malaria will have different symptoms. Flu and malaria will both cause fever. Flu will also cause aches. Malaria will also cause jaundice. Having a fever will also cause me to have high temperature.

For each question, I will give you the answer first. then, I will discuss how I derived the answer. Make sure that you pause the video and derive the answer yourself before looking at my explanations.



Problem: **(1a)** Are TravelSubway and HighTemp independent?

Solution:

The correct answer is NO. TravelSubway and HighTemp are not unconditionally independent.

First, find all the paths between the two nodes. There is only one path between the two nodes: TravelSubway - Flu - Fever - HighTemp.

Next, consider every node on the path and verify whether the node blocks the path. Flu is not observed and does not block the path by rule 1. Fever is not observed and does not block the path by rule 1. Since the path is not blocked, TravelSubway and HighTemp are NOT independent.

Problem: **(1b)** Are TravelSubway and HighTemp conditionally independent given Flu?

Solution:

The correct answer is YES. TravelSubway and HighTemp are not independent given Flu.

Consider the same path again. Flu is observed and it does block the path by rule 1. As soon as one node blocks the path, we can stop checking. For your information, Fever is not observed and it does not block the path by rule 1. Since the path is blocked, TravelSubway are HighTemp are independent given Flu.

Problem: **(2a)** Are Aches and HighTemp independent?

Solution:

The correct answer is NO. Aches and HighTemp are NOT independent.

We need to check one path Aches - Flu - Fever - HighTemp. Flu is not observed and does not block the path by rule 2. Fever is not observed and does not block the path by rule 1. Since the path is not blocked, Aches and HighTemp are NOT independent.

Problem: **(2b)** Are Aches and HighTemp conditionally independent given Flu?

Solution:

The correct answer is YES. Aches and HighTemp are independent given Flu.

We need to check one path Aches - Flu - Fever - HighTemp. Flu is observed and blocks the path by rule 2. As soon as one node blocks the path, we can stop checking. For your information, Fever is not observed and it does not block the path by rule 1. Since the path is blocked, Aches and HighTemp are independent given Flu.

Problem: **(3a)** Are Flu and ExoticTrip independent?

Solution:

The correct answer is YES. Flu and ExoticTrip are independent.

We need to check one path: Flu - Fever - Malaria - ExoticTrip. Fever is not observed and its descendants are not observed. Thus, Fever blocks the path by rule 3. As soon as one node blocks the path, we can stop checking. For your information, Malaria is not observed and it does not block the path by rule 1. Since the path is blocked, Flu and ExoticTrip are independent.

Problem: **(3b)** Are Flu and ExoticTrip conditionally independent given HighTemp?

Solution:

The correct answer is NO. Flu and ExoticTrip are NOT independent given HighTemp.

We need to check one path: Flu - Fever - Malaria - ExoticTrip. Fever is not observed but Fever's child HighTemp is observed. Thus, Fever blocks the path by rule 3. Malaria is not observed and it does not block the path by rule 1. Since the path is not blocked, Flu and ExoticTrip are NOT independent given HighTemp.

2 Constructing Bayesian Networks

2.1 Multiple Correct Bayesian Networks

I have introduced the Holmes story and showed you a Bayesian network that can capture this story. This is not the only network for this story. We can construct many different Bayesian networks, and all of them are valid models of this story. In general, for any joint probability distribution, we can construct multiple correct Bayesian networks.

What does it mean for a Bayesian network to be correct? A Bayesian network is correct if the following condition is satisfied:

"If the Bayesian network requires the variables to satisfy an unconditional or conditional independence relationship, the joint distribution must also require the variables to satisfy the same independence relationship."

In many scenarios, we already have a Bayesian network and we want to construct another Bayesian network that represents the same scenario. In this case, a correct Bayesian network must satisfy the following condition:

"If the new Bayesian network requires the variables to satisfy an unconditional or conditional independence relationship, the old Bayesian network must also require the variables to satisfy the same independence relationship."

This is a conditional statement, and its contrapositive must also be true as well. So, the contrapositive is:

"If the old Bayesian network does not require two variables to satisfy an independence relationship, then the new Bayesian network also cannot require the two variables to satisfy the independence relationship."

If the old network has an edge between A and B, then the new network must also have an edge between A and B. However, if the old network does not have an edge between A and B, there is no requirement on the relationship between A and B in the new network.

If there are multiple correct Bayesian networks for each distribution, which one should we choose? Our original reason for using a Bayesian network is to reduce the number of probabilities required to represent the distribution. Therefore, in general, we prefer a Bayesian network that requires us to specify fewer probabilities. The number of probabilities required roughly corresponds to the number of edges in the network. Thus, roughly speaking, we prefer a network with fewer edges. The smaller the Bayesian network, the easier to work with it.

2.2 Requiring an Independence Relationship

To understand the definition of a correct Bayesian network, we need to understand a key point: What does it mean for a Bayesian network to require an independence relationship? Let's look at two cases.

First case: if an edge connects two variables, it does not mean that the two variables are dependent on each other. The presence of an edge does not indicate that there is an dependence relationship between the two variables.

Take a look at the simple example. There is an edge from A to B. This network could represent a distribution in which A and B are not independent. Here is an example. P(B—A) is not equal to P(B), so B and A re not independent.

This network could also represent a distribution in which A and B are independent. Here is an example. Note that P(B—A) is equal to P(B).

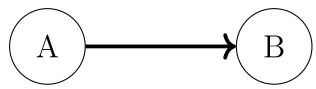


Figure 4: First Case

Second case: If there is no edge between the two variables, it means that the two variables must satisfy an independence relationship. The absence of an edge means that the Bayesian network requires the two variables to satisfy an independence relationship.

This is a simple example. There are two nodes and there is no edge between them. This network requires A and B to be unconditionally independent. This network can represent the distribution in which A and B are independent. It cannot represent the distribution in which A and B are not independent.

Figure 5: Second Case

2.3 A Procedure for Constructing A Correct Bayesian Network

Let's look at a procedure to construct a correct Bayesian network.

Example:

1. Order the variables $\{X_1, \ldots, X_n\}$.
2. For each variable X_i in the ordering:

   2.1 Choose the node's parents.

   Choose the smallest set of parents from $\{X_1, \ldots, X_{i-1}\}$ such that given Parents(X_i), X_i is independent of all the nodes in $\{X_1, \ldots, X_{i-1}\}$ − Parents(X_i). Formally,

$$P(X_i|\text{Parents}(X_i)) = P(X_i|X_{i-1} \wedge \cdots \wedge X_1).$$

2.2 Create a link from each parent of X_i to the node X_i.

2.3 Write down the conditional probability table $P(X_i|\text{Parents}(X_i))$.

Step 1: We are given a joint distribution involving several variables. Alternatively, we are given a Bayesian network that captures the joint distribution correctly.

Step 2: Order the variables in some way. This order will significantly affect the structure of the resulting Bayesian network.

Step 3: We will construct the new Bayesian network by adding the variables one by one based on the chosen order.

Whenever we add a variable to the network, we must determine the parent nodes of this new variable. In other words, we may need to connect the new variable to some existing variables. Once we choose the parent nodes, we will create one directed edge from each parent to the new node, and we will specify the conditional probability table at this new node.

The most challenging step is 3.1. How do we choose the new node's parents? This is a rare opportunity where we get to choose our parents, so we should do so carefully.

In general, we can always make all the existing nodes to be parents of the new node. This will result in a fully connected Bayesian network. However, this is not a great strategy. Our goal of using a Bayesian network is to reduce the number of probabilities that we need to specify.

The rule is: we need to choose the smallest set of parent nodes such that the parent nodes make the new node independent of all the other existing nodes.

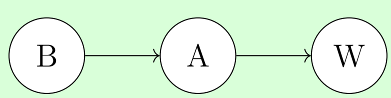
In an extreme case, the set of parents could be empty. This means that the new node does not connect to any existing node.

2.4 Examples

2.4.1 Example 1

Problem:

Consider the Bayesian network:



Construct a correct Bayesian network based on the variable ordering: W, A, B.

Solution:

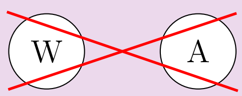
Step 1: add W to the network.

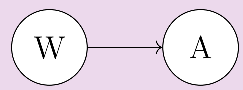
Step 2: add A to the network.

What are the parent nodes of A? The network had one node before adding A. We have two options: Either A has no parent, or W is A's parent.

Recall the rule for choosing the parent nodes. We need to choose the smallest set of parents that make the new node independent of all the other existing nodes.

In the original network, there is a directed edge from A to W. The original network does not require A and W to be independent. Given this, the new network also cannot require A and W to be independent. Therefore, there must be an edge from W to A.





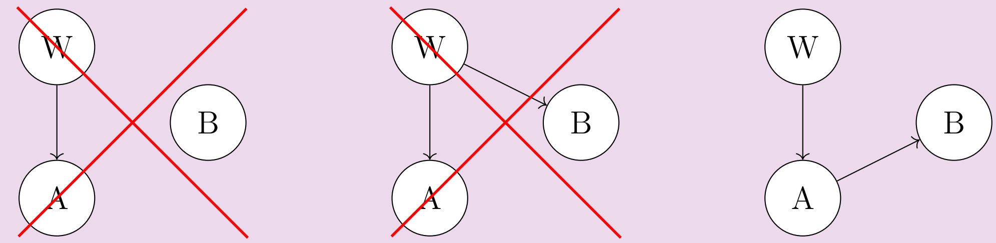
Step 3: add B to the network.

What are the parents of B? There are two nodes in the network before adding B. We have four options: no parent, W is the only parent, A is the only parent, and W and A are both parents. We can consider these four options in any order. Let me show you how I would tackle this.

There is an edge from B to A in the original network. Given this edge, the original network does not require A and B to be independent. Thus, the new network cannot require A and B to be independent as well. So A must be B's parent.

Does W need to be B's parent? In the original network, there is no edge between B and W. The original network requires B and W to be independent given A. Based on the definition of a correct network, this means that there is no requirement on the relationship between B and W in the new network. The new network is correct regardless of whether W is B's parent or not.

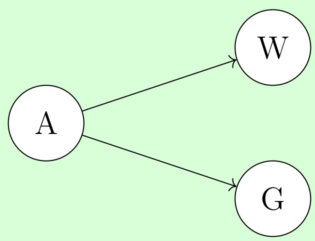
We have two correct options remain: A is the only parent of B. W and A are both parents of B. Since we want to choose the smallest set of parents, we will choose the first option: A is B's only parent.



2.4.2 Example 2

Problem:

Consider the Bayesian network:



Construct a correct Bayesian network based on the variable ordering: W, G, A.

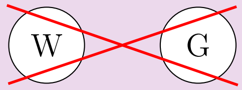
Solution: Step 1: add W to the network.

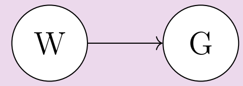
Step 2: add G to the network.

What are the parent nodes of G? The network had one node before adding G. We have two options: Either G has no parent, or W is G's parent.

If W is not G's parent, then the new network requires G and W to be unconditionally independent (since we haven't added A to the network yet). Can the new network require this?

Let's look at the original network. In the original network, there is no edge between W and G. By d-separation, W and G are independent given A. However, if we haven't observed A, W and G are not unconditionally independent. Since the original network does not require W and G to be unconditionally independent, the new network also cannot require W and G to be unconditionally independent. Therefore, W must be G's parent.



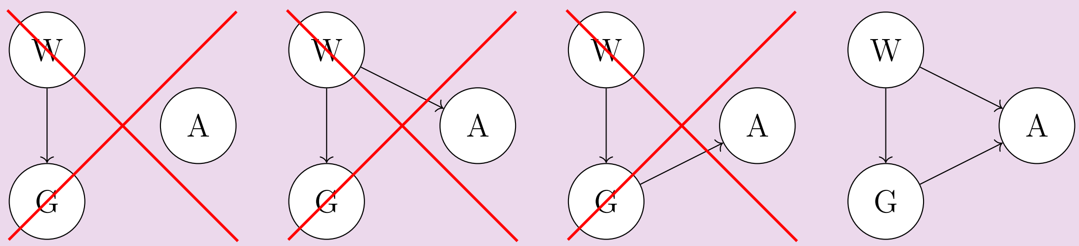


Step 3: add A to the network.

What are the parent nodes of A? W and G were added before adding A. We have four options: no parent, W is the only parent, G is the only parent, and W and G are both parents of A.

In the original network, W and A are directly connected. The original network does not require W and A to be independent. Therefore, the new network cannot require W and A to be independent. W must be A's parent.

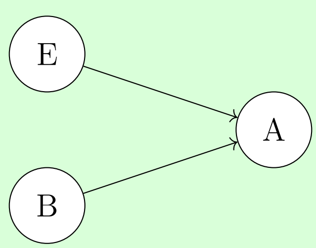
Similarly, in the original network, G and A are directly connected. The original network does not require G and A to be independent. Therefore, the new network cannot require G and A to be independent. G must be A's parent as well.



2.4.3 Example 3

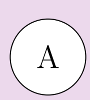
Problem:

Consider the Bayesian network:



Construct a correct Bayesian network based on the variable ordering: A, B, E.

Solution: Step 1: add A to the network.



Step 2: add B to the network.

What are the parent nodes of B? The network had one node before adding B. We have two options: Either B has no parent, or A is B's parent.

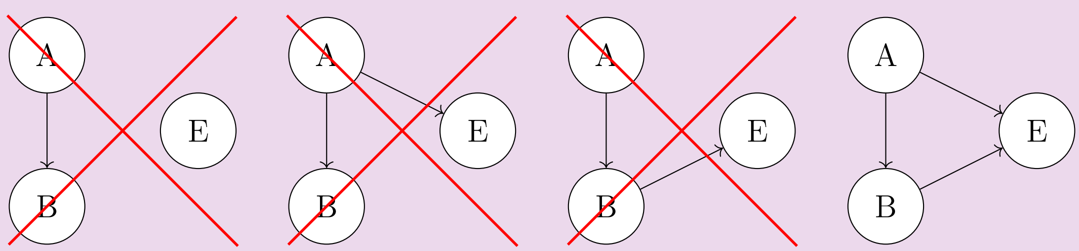
In the original network, A and B are directly connected. The original network does not require A and B to be independent. Thus, the new network cannot require A and B to be independent. A must be B's parent.

Step 3: add E to the network.

What are the parent nodes of E? A and B were added before E. We have four options: no parent, A is the only parent, B is the only parent, and A and B are both parents of E.

In the original network, A and E are directly connected. The original network does not require A and E to be independent. Therefore, the new network cannot require A and E to be independent. A must be E's parent.

Should B be E's parent or not? If B is not E's parent, then the new network requires E to be independent of B given A. Can the new network require this? Let's look at the original network. In the original network, B and E are not independent given A. Since the original network does not require B and E to be independent given A, then the new network also cannot require B and E to be independent given A. Therefore, B must also be E's parent.



2.5 Discussion

Let me discuss two important points.

First, what does each edge represent? Some edges may appear quite unintuitive to you. Why is there an edge from Burglary to Earthquake? Burglary certainly does not cause Earthquake to happen.

Your intuition is correct here. An edge in a Bayesian network does not necessarily represent a causal relationship. An edge only represents correlation, not causality.

The Bayesian network for the Holmes story was constructed by adding nodes to the network in a causal order. That's why the edges represent causal relationships. This is not the case in general.

Second, I mentioned earlier that we prefer a network with fewer edges. How can we construct a network with the smallest number of edges? By examples 2 and 3, you have observed that the order of adding the variables significantly affects the number of edges in the network. To construct a network with fewer edges, try adding the variables in a causal order. For variables representing causes and effects, add the causes to the network before adding the effects. For example, if A causes B, add A before adding B. If we add the variables following a causal order, we tend to get a Bayesian network with fewer edges.