Lecture 2
Conditional Independence and d-Separation

Course Probabilistic Graphical Models, September 12th, 2017

Johannes Textor
Institute for Computing and Information Sciences
Learning Objectives 2

1. Understand the concept of conditional independence.
2. Understand and be able to apply the $d$-separation criterion.
3. Be able to test Bayesian network structure against data.
Lecture Outline 2

1. Conditional Independence

2. $d$-Separation

3. Testing conditional independence
Motivation

- Bayes nets are **models** of variable relationships in a certain domain.
- Sparse Bayes net models encode certain **assumptions** about these relationships.
- Incorrect assumptions may lead to **incorrect** inferences.
- Once a Bayes net is constructed, we can **test** some of the assumptions it encodes against data.

No free lunch!

Model testing never guarantees a correct model! It can only refute, but never prove it.
Conditional Independence

Two variables $X$, $Y$ are called independent if

$$P(x, y) = P(x)P(y).$$

Two variables $X$, $Y$ are called conditionally independent given a set of variables $Z$ if

$$P(x, y | z) = P(x | z)P(y | z).$$

Equivalently, $X$ and $Y$ are called independent given $Z$ if

$$P(x | y, z) = P(x | z).$$

Interpretation: Once we know $Z$, $Y$ provides us no additional information about $X$.

Notation

- $X \perp Y$ means: $X$ and $Y$ are independent.
- $X \perp Y | Z$ means: $X$ and $Y$ are independent given $Z$. 
We call a probability density $P$ consistent with $G$ if $P$ factorizes according to $G$.

It appears that consistency is intimately linked to statistical dependencies.
Example: The Mediation Model

Take the following Bayes net, known as the “full mediation model”:

\[ X \rightarrow M \rightarrow Y \]

The factorization is:

\[ P(x, m, y) = P(x)P(m \mid x)P(y \mid m) = P(x \mid m)P(y \mid m)P(m) \]

Therefore,

\[ P(x, y \mid m) = \frac{P(x, y, m)}{P(m)} = P(x \mid m)P(y \mid m) \]

and this means that \( X \perp Y \mid M \).
## 5-Minute Exercise

### Exercise

The Bayes net $X \rightarrow M \rightarrow Y$ "claims" that $X$ and $Y$ are conditionally independent given $\{M\}$. Fill in the missing probabilities into the table below such that this claim is violated.

<table>
<thead>
<tr>
<th>$M$</th>
<th>$X$</th>
<th>$Y$</th>
<th>$P$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1/8</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1/8</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1/8</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1/8</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1/8</td>
</tr>
</tbody>
</table>
In the mediation example, we derived a conditional independence by re-arranging the factorization. This always works, but it is tedious. 

*d*-Separation is a **graphical criterion** that allows us to derive all conditional independencies from the Bayes net graph.

**Logic of the *d*-separation criterion**

Let $G$ be a DAG and let $P$ be a probability density that factorizes according to $G$.

- If $X$ and $Y$ are “*d*-separated” by $Z$, then $X \perp Y \mid Z$ is guaranteed to hold.
- If $X$ and $Y$ are not “*d*-separated” by $Z$, then $X \perp Y \mid Z$ may or may not to hold.

In other words, *d*-separation is a **sufficient**, but not a **necessary** criterion.
A path in a Bayes net is a sequence of variables in which each adjacent pair is connected by an edge.

Note that this differs from the classic graph-theoretical concept of a path: it is also allowed to move against arrow directions. $X \rightarrow M \leftarrow Y$ is a path in the Bayes-net sense, but not in the classic sense.

By convention, we consider only those paths that contain each variable at most once.
The 3-Variable Case

Let us consider the simplest four Bayes nets with two unconnected variables.

<table>
<thead>
<tr>
<th>Network</th>
<th>Name</th>
<th>Independence implication</th>
</tr>
</thead>
<tbody>
<tr>
<td>$X \rightarrow M \rightarrow Y$</td>
<td>chain</td>
<td>Unconditional: none, Conditional: $X \perp Y \mid {M}$</td>
</tr>
<tr>
<td>$X \leftarrow M \leftarrow Y$</td>
<td>inverse chain</td>
<td>Unconditional: none, Conditional: $X \perp Y \mid {M}$</td>
</tr>
<tr>
<td>$X \leftarrow M \rightarrow Y$</td>
<td>fork</td>
<td>Unconditional: none, Conditional: $X \perp Y \mid {M}$</td>
</tr>
<tr>
<td>$X \rightarrow M \leftarrow Y$</td>
<td>collider</td>
<td>Unconditional: $X \perp Y$, Conditional: none</td>
</tr>
</tbody>
</table>

For a path $\pi = (X, M, Y)$, we define:

1. If $\pi$ is a collider $X \rightarrow M \leftarrow Y$:
   - $\{\}$ d-separates $X$ and $Y$
   - $\{M\}$ d-connects $X$ and $Y$

2. If $\pi$ is not a collider:
   - $\{\}$ d-connects $X$ and $Y$
   - $\{M\}$ d-separates $X$ and $Y$
Example

There is no striking correlation between IQ and wealth in the general population. But surveying students on a campus of a private elite university, one might well find a striking inverse correlation – smarter students tend to be poorer. Why could this happen?

Pay tuition (wealth)  
Get scholarship (IQ)  
Go to university
Consider a Bayes net that consists of a single path $\pi = (X_1, X_2, \ldots, X_n)$, and a set $Z \subseteq \{X_2, \ldots, X_{n-1}\}$. We say that $Z$ $d$-separates $X_1$ and $X_n$ if:

- $\pi$ contains a collider $X_{i-1} \rightarrow X_i \leftarrow X_{i+1}$, where $X_i \notin Z$;
- $\pi$ contains a non-collider
  - $X_{i-1} \rightarrow X_i \rightarrow X_{i+1}$,
  - $X_{i-1} \leftarrow X_i \rightarrow X_{i+1}$, or
  - $X_{i-1} \leftarrow X_i \leftarrow X_{i+1}$,

where $X_i \in Z$. 
5-Minute Exercise

Exercise

Take the Bayes net below.

- Give two sets $Z$ that $d$-separate $A$ and $F$.
- Give a set $Z$ containing $D$ that $d$-separates $A$ and $F$. 
An additional rule is required for colliders that appear in larger graphs. In this Bayes net, we know that conditioning on $M$ can render $X$ and $Y$ dependent.

$$X \rightarrow M \leftarrow Y$$

$$\downarrow$$

$$D$$

But: conditioning on $D$ can also render $X$ and $Y$ dependent. To see this, consider that $D$ and $M$ could be very strongly dependent themselves. Then conditioning on $D$ and conditioning on $M$ would have very similar effects.
The $d$-Separation Criterion

Consider a Bayes net $G$ with variables $V = \{V_1, \ldots, V_n\}$. We say that $Z \subseteq V \setminus \{V_i, V_j\}$ $d$-separates $V_i$ and $V_j$ in $G$ if for every path $\pi = (V_i, V_{k_1}, \ldots, V_{k_n}, V_j), n \geq 0$,

- $\pi$ contains a collider $X_{i-1} \rightarrow X_i \leftarrow X_{i+1}$, such that $X_i$ is not an ancestor of any node in $Z$; or
- $\pi$ contains a non-collider
  - $X_{i-1} \rightarrow X_i \rightarrow X_{i+1}$,
  - $X_{i-1} \leftarrow X_i \rightarrow X_{i+1}$, or
  - $X_{i-1} \leftarrow X_i \leftarrow X_{i+1}$,

where $X_i \in Z$.

Theorem (Verma & Pearl, 1984)

*If $Z$ $d$-separates $X$ and $Y$ in a DAG $G$, then $X \perp Y \mid Z$ in every probability density $P$ that factorizes according to $G$.***
The Ancestral Graph

The $d$ in $d$-separation stands for “directed”. But surprisingly, the somewhat intricate $d$-separation criterion can be quite elegantly reduced to standard separation in undirected graphs (à la max-flow-min-cut). We need two definitions for this.

**Definition**

Given a DAG $G = (V, E)$ and a variable subset $Z \subseteq V$, the **ancestral graph** $G_a(Z)$ is obtained as follows: Delete all variables except those that are ancestors of any variable in $Z$. 

![Diagram of the ancestral graph](image)

\[ G = \]

\[ G_a(\{H, B\}) = \]

\[ G = \]

\[ G_a(\{H, B\}) = \]
The Moral Graph

Definition

Given a DAG $D$, the moral graph $G^m$ is the undirected graph created as follows:

1. Connect all variables that have common children with an undirected edge.
2. Replace all directed by undirected edges between the same nodes.

$$G = \begin{align*}
    & A & R \\
    T & H & \leftarrow L & \rightarrow B
\end{align*}$$

$$G^m = \begin{align*}
    & A & R \\
    T & H & \leftarrow L & \rightarrow B
\end{align*}$$

Theorem (Lauritzen & Spiegelhalter)

A set $Z$ d-separates $X$ and $Y$ in the DAG $G$ if and only if $Z$ separates $X$ and $Y$ in the ancestor moral graph $(G_a(\{X, Y\} \cup Z))^m$. 
Failed $d$-separation implications inform us about errors in the model structure. For instance, in the example below, our Bayes net fails to take into account an important variable $U_1$. How could we detect such a mistake?

\[
\begin{align*}
\text{assumed model} & \quad M_1 \perp M_2 \mid X \\
X \perp Y \mid \{M_1, M_2\}
\end{align*}
\]

\[
\begin{align*}
\text{true model} & \quad M_1 \perp M_2 \mid X \\
X \perp Y \mid \{M_1, M_2\}
\end{align*}
\]
Generating Data from True Model

Let us generate some binary data that follows the “true” Bayes net structure. We use a small simulation in R for this.

```r
set.seed(123)
# A utility function that converts log-odds to probabilities.
odds2p <- function(o) exp(o)/(exp(o)+1)
n <- 10000

# X does not depend on anything.
X <- rbinom(n, 1, .5)
# X=1 increases the odds of U1=1 and M2=1.
U1 <- rbinom(n, 1, odds2p(4*X-2))
M2 <- rbinom(n, 1, odds2p(4*X-2))
# U1=1 increases the odds of M1=1.
M1 <- rbinom(n, 1, odds2p(2*U1-1))
# U1=1 and M2=1 both increase the odds of Y=1.
Y <- rbinom(n, 1, odds2p(2*U1+2*M2))
```

```
X -> U1 -> M1
X -> M2 -> Y
```
Testing the Assumed Model (I)

Let us start with the first implied independence: $M_1 \perp M_2 \mid X$. A general strategy for testing this is performing standard independence tests between $M_1$ and $M_2$ in every stratum (for every value) of $X$, and combining the results.

```r
chisq.test( M1[X==0], M2[X==0] )
```

## Pearson's Chi-squared test with Yates' continuity correction
## data: M1[X == 0] and M2[X == 0]
## X-squared = 0.66084, df = 1, p-value = 0.4163

```r
chisq.test( M1[X==1], M2[X==1] )
```

## Pearson's Chi-squared test with Yates' continuity correction
## data: M1[X == 1] and M2[X == 1]
## X-squared = 1.0784, df = 1, p-value = 0.2991

There is no evidence for dependence here.
Testing the Assumed Model (II)

Now let us use the same approach to test the second implied independence: $X \independent Y \mid \{M_1, M_2\}$.

```r
chisq.test( X[M1==1 & M2==1], Y[M1==1 & M2==1] )
```

```r
## Pearson's Chi-squared test with Yates' continuity correction
## data: X[M1 == 1 & M2 == 1] and Y[M1 == 1 & M2 == 1]
## X-squared = 53.846, df = 1, p-value = 2.169e-13
```

If a dependence can be found for any stratum of the conditioning variables, the conditional independence is refuted.

Note

Obviously, for large strata, we will have to perform a large number of tests. This gives rise to both computational (runtime) and statistical issues (multiple testing problem). Those issues will not be covered in detail today.
Summary of Test Results

- We assumed the following Bayes net:
  \[
  \begin{align*}
  &M_1 \\
  X &\rightarrow M_1 \rightarrow Y \\
  &M_2
  \end{align*}
  \]

- Using \(d\)-Separation, we derived two conditional independencies from the net:
  1. \(M_1 \perp M_2 \mid X\)
  2. \(X \perp Y \mid \{M_1, M_2\}\)

- We could not refute the first independence.
- We did refute the second independence.
Summary

1. Bayesian networks put conditional independence constraints on compatible probability distributions.
2. The $d$-separation criterion allows to read off these constraints from the graphical model structure.
3. The constraints can be tested parametrically.