

Lecture 2: April 5, 2002

Lecturer: Jeff Bilmes

Scribe: Chris Bartels/Scott Otterson

2.1 Directed Graphical Models

We will be using the following graph to illustrate several concepts over the next several lectures:

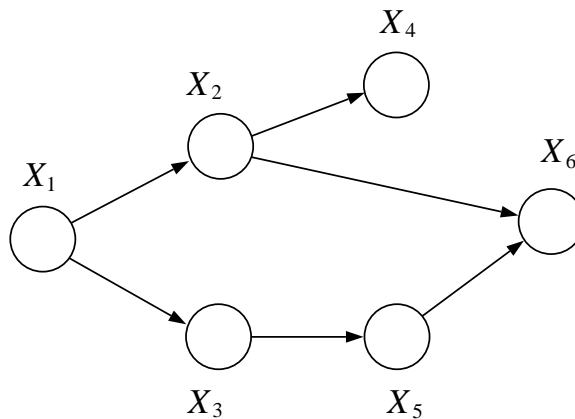


Figure 2.1: Example of a six node GM

From last lecture, we know the joint probability of $x_{1:n}$ according to a graph is defined by the following factorization:

$$p(x_{1:n}) \triangleq \prod_i p(x_i | x_{\pi_i})$$

In our example graph (figure 2.1), this factorization gives:

$$p(x_{1:6}) = p(x_1)p(x_2|x_1)p(x_3|x_1)p(x_4|x_2)p(x_5|x_3)p(x_6|x_2, x_5)$$

If X is a discrete variable with cardinality r , the joint probability of an arbitrary 6-variate distribution can be represented in a table with r^6 entries. The factorization of our GM can be stored in a table with only r^3 entries. This reduction from $O(r^6)$ to $O(r^3)$ will result in huge savings for computation and storage.

More generally if:

⁰These notes are partially based on those of Daniel Gatica and Scott Otterson.

$$m_i \triangleq \# \text{ of parents of node } i$$

$$|X_i| = r$$

$p(x_i|x_{\pi_i}) \rightarrow$ can be represented in table of size r^{m_i+1} , giving savings of:

$$O(r^n) \rightarrow O(r^{\max_i m_i + 1})$$

In addition to this computational savings, the GM makes it easier to understand and discuss a particular distribution.

2.2 Independence

X_A is (marginally) independent of $X_B \equiv X_A \perp\!\!\!\perp X_B$. We will give two equivalent definitions:

- $X_A \perp\!\!\!\perp X_B$ if:

$$p(x_A, x_B) = p(x_A)p(x_B) \quad \forall x_A, x_B$$

- or equivalently, $X_A \perp\!\!\!\perp X_B$ if and only if:

$$\text{There exists functions } g \text{ and } h \text{ s.t. } p(x_A, x_B) = g(x_A)h(x_B) \quad \forall x_A, x_B$$

2.3 Conditional Independence

X_A is conditionally independent of X_B given $X_C \equiv X_A \perp\!\!\!\perp X_B | X_C$. We will give four equivalent definitions.

- $X_A \perp\!\!\!\perp X_B | X_C$ if:

$$p(x_A, x_B | x_C) = p(x_A | x_C)p(x_B | x_C) \quad \forall x_A, x_B, x_C$$

- equivalently, $X_A \perp\!\!\!\perp X_B | X_C$ if:

$$p(x_A | x_B, x_C) = p(x_A | x_C) \quad \forall x_A, x_B, x_C$$

This can be easily proven using the chain rule.

- equivalently, $X_A \perp\!\!\!\perp X_B | X_C$ if:

$$p(x_A, x_B, x_C) = \frac{p(x_A, x_C)p(x_B, x_C)}{p(x_C)} \quad \forall x_A, x_B, x_C$$

This definition represents conditional independence as a junction tree. More to come. . .

- equivalently, $X_A \perp\!\!\!\perp X_B | X_C$ if and only if:

$$\text{There exists functions } g \text{ and } h \text{ s.t. } p(x_A, x_B, x_C) = g(x_A, x_C)h(x_B, x_C) \quad \forall x_A, x_B, x_C$$

Node Subsets: We will show $X_A \perp\!\!\!\perp X_B | X_C$ implies $X_{A'} \perp\!\!\!\perp X_{B'} | X_C$ when $A' \subseteq A$ and $B' \subseteq B$:

$$\int_{A \setminus A', B \setminus B'} p(X_A, X_B | X_C) dX_{A,B,C} = \int_{A \setminus A', B \setminus B'} p(X_A, X_C) p(X_B, X_C) dX_{A,B,C}$$

$$p(x_{A'}, x_{B'} | x_C) = p(x_{A'} | x_C) p(x_{B'} | x_C)$$

So, subsets inherit the conditional independencies of their supersets.

Example:

$$P(X, Y, Z, A, B, C, D, E) = P(Y|D)P(D|C, E)P(E|Z)P(C|B)P(B|A)P(X|A)$$

Is $X \perp\!\!\!\perp Z | Y$? \rightarrow This is difficult to determine from the factorization.

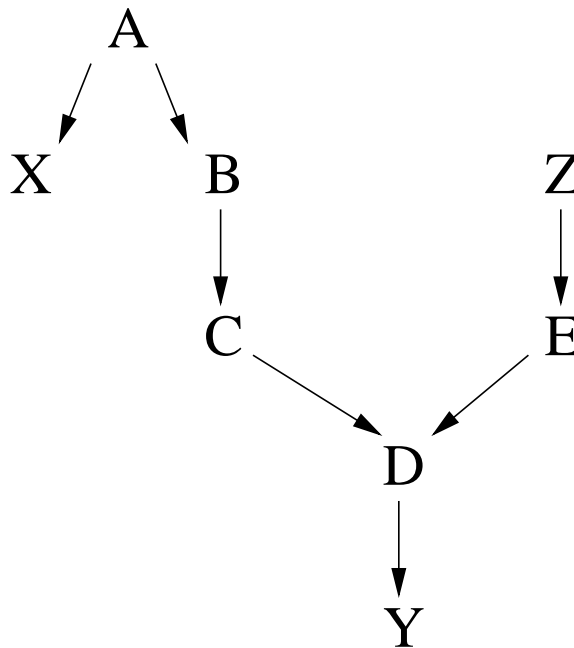


Figure 2.2: Graph corresponding to the factorization

Again, is $X \perp\!\!\!\perp Z | Y$? \rightarrow No! This is easy to determine from the graph.

Note that **missing variables** in local conditional pdfs correspond to **missing edges** in the GM. This can be seen in the following expression, obtained from the six-node example:

$$p(x_4 | x_{1:3}) = p(x_4 | x_2) \rightarrow x_4 \perp\!\!\!\perp \{x_1, x_3\} | x_2$$

2.4 Semantics of a DGM

Definition 2.1. A **topological (total) ordering** I of the graph G is such that all parents of node i occur earlier in I than i .

In the six-node example:

- $I = \{1, 2, 3, 4, 5, 6\}$ is a total ordering.
- $I = \{6, 2, 5, 3, 1, 4\}$ is *not* a total ordering.

Definition 2.2. We define the set v_i to be the set of indices before i in I other than π_i .

$$v_i = \{j < i : j \in I, j \notin \pi_i\}$$

In the six node example:

$$I = \{1, 2, 3, 4, 5, 6\}$$

$$v = \{\emptyset, \emptyset, \{2\}, \{1, 3\}, \{1, 2, 4\}, \{1, 3, 4\}\}$$

Given topological ordering I , the graph G makes the following conditional independence statements:

$$X_i \perp\!\!\!\perp X_{v_i} \mid X_{\pi_i}$$

This is a "Markov Property" on directed graphs:

$$X_i \perp\!\!\!\perp nd(X_i) \mid pa(X_i)$$

Where $nd(\cdot)$ denotes non-descendants, and $pa(\cdot)$ denotes parents. This means that x is independent of its grandparents given its parents, but it is not independent of its children.

Example.

In the six-node example, $I = \{1, 2, 3, 4, 5, 6\}$, and $v = \{\emptyset, \emptyset, \{2\}, \{1, 3\}, \{1, 2, 4\}, \{1, 3, 4\}\}$. We can use these to make the following independence statements:

- $X_1 \perp\!\!\!\perp \emptyset \mid \emptyset$.
- $X_2 \perp\!\!\!\perp \emptyset \mid X_1$.
- $X_3 \perp\!\!\!\perp X_2 \mid X_1$.
- $X_4 \perp\!\!\!\perp \{X_1, X_3\} \mid X_2$.
- $X_5 \perp\!\!\!\perp \{X_1, X_2, X_4\} \mid X_3$.
- $X_6 \perp\!\!\!\perp \{X_1, X_3, X_4\} \mid \{X_2, X_5\}$.
- There exists additional independence statements

We can also derive conditional independence statements from the factorization using algebra. Let's illustrate this by an example.

Example. Show $X_4 \perp\!\!\!\perp X_{1,3} \mid X_2$:

Solution. By definition,

$$\begin{aligned}
 p(x_{1:4}) &= \sum_{x_{5:6}} p(x_{1:6}) \\
 &= p(x_1)p(x_2|x_1)p(x_3|x_1)p(x_4|x_2) \sum_{x_{5:6}} p(x_5|x_3)p(x_6|x_2, x_5) \\
 &= p(x_1)p(x_2|x_1)p(x_3|x_1)p(x_4|x_2) \sum_{x_5} p(x_5|x_3) \sum_{x_6} p(x_6|x_2, x_5) \\
 &= p(x_1)p(x_2|x_1)p(x_3|x_1)p(x_4|x_2)
 \end{aligned}$$

Similarly, it can be shown that

$$p(x_{1:3}) = \sum_{x_4} p(x_{1:4}) = p(x_1)p(x_2|x_1)p(x_3|x_1)$$

And therefore

$$p(x_4|x_{1:3}) = \frac{p(x_{1:4})}{p(x_{1:3})} = p(x_4|x_2)$$

and from here we can conclude that $X_4 \perp\!\!\!\perp \{X_1, X_3\} | X_2$.

2.5 Three canonical GM's

Case I. Serial Connection (Markov Chain)

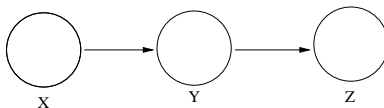


Figure 2.3:

From this graph we can conclude:

$$\begin{aligned}
 X &\perp\!\!\!\perp Z | Y \\
 I &= \{X, Y, Z\} \\
 v &= \{\phi, \phi, \{X\}\}
 \end{aligned}$$

We can see this as if Y had blocked the path between X and Z (node Y could be shaded). Why is the CI assertion true?

$$p(x, y, z) = p(x)p(y|x)p(z|y)$$

Therefore

$$p(z|x, y) = \frac{p(x, y, z)}{p(x, y)} = \frac{p(x)p(y|x)p(z|y)}{p(x)p(y|x)} = p(z|y)$$

No other CI statements characterize this graph *in general*. In particular, we could choose x and y such that $p(y|x) = p(y)$, but the model says in general, only that $X \perp\!\!\!\perp Z | Y$. As a general result, we can say that

- Edges that are present do not necessarily imply dependence. Edges allow for a dependence.
- Edges that are missing necessarily imply conditional independence.
 $A \rightarrow B$ describes every possible distribution.
 $A \rightarrow B \rightarrow C$ has a missing edge and constrains the possible distributions.
- Asserted CI always hold for distributions that correspond to the graph.
- Non-asserted CI (i.e. $X \perp\!\!\!\perp Y$) sometimes hold and other times fail to hold.

Case II. Diverging Connection

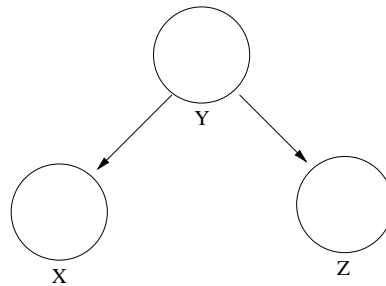


Figure 2.4:

From this graph we can conclude:

$$X \perp\!\!\!\perp Z | Y$$

$$I = \{Y, X, Z\}$$

$$V = \{\phi, \phi, \{X\}\}$$

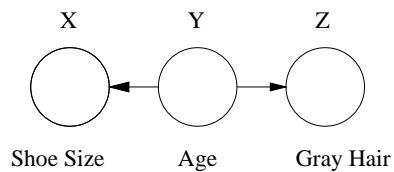


Figure 2.5: Example of a diverging connection

Age is a hidden confounding cause. Shoe size does not cause grey hair.

Case III. Converging Connection

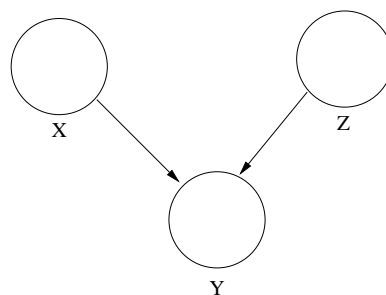


Figure 2.6:

From this graph we can conclude:

$$I = \{X, Z, Y\}$$

$$V = \{\phi, \{X\}, \phi\}$$

$$X \perp\!\!\!\perp Z$$

$$p(x, z) = \sum_y p(x, y, z) = \sum_y p(y|x, z)p(x)p(z) = p(x)p(z)$$

Is $X \perp\!\!\!\perp Z | Y \leftarrow$ No!

Figure 2.6 could describe:

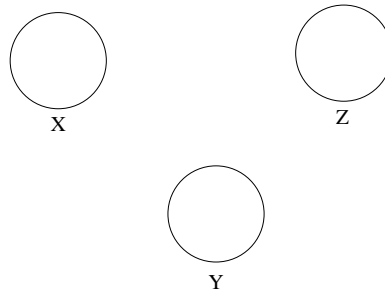


Figure 2.7:

Where X , Y , and Z are all independent. This is because the second graph has fewer independence statements, so it represents a larger family of distributions.

Examples of converging connections:

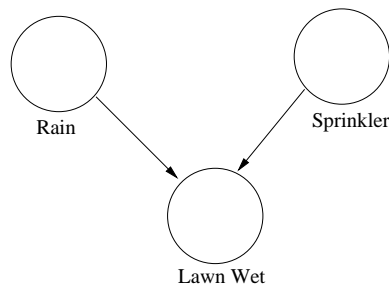


Figure 2.8:

Rain and a sprinkler are both explanations for the lawn being wet. The fact that it is raining does not affect whether the sprinkler is turned on or off.

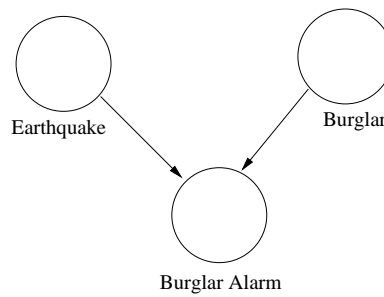


Figure 2.9:

An earthquake and a burglar are both reasons that the burglar alarm could be activated. The activities of burglars does not change the chances of an earthquake happening.

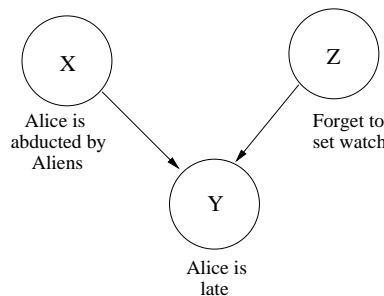


Figure 2.10:

Let's assume a converging connection GM, and let X , Y , and Z denote "Alice is abducted by aliens", "Alice is late", and "Bob forgot to set his watch". In this example, we can assert that

$$P(X = \text{yes} | Y = \text{yes}) > P(X = \text{yes})$$

and

$$P(X = \text{yes} | Y = \text{yes}, Z = \text{yes}) \leq P(X = \text{yes} | Y = \text{yes})$$

Therefore

$$p(x|y, z) \neq p(x|y) \rightarrow X \not\perp\!\!\!\perp Z | Y \text{ does not hold}$$

This is the notion of explaining away.

D-Separation: In Case III, $X \not\perp\!\!\!\perp Z | Y$ even though in the graph X is "separated" from Z by Y . In directed graphs, we will see that we need a more sophisticated notion of graph separation if we want it to correspond to conditional independence. One way of doing this is with d-separation, which stands for directed-separation, and is our second Markov property.

Definition of d-separation (for directed graphs). Two sets of nodes (A, B) are d-separated by node set C if \forall paths between a node in A and a node in B , $\exists v$ in the path such that either:

1. The connection is either serial at v or diverging at v and $v \in C$ (i.e. v is shaded or on the right hand side of the conditioning bar).
2. The connection is converging or neither v nor any of v 's descendants live in C .

Theorem 2.3. $X_A \perp\!\!\!\perp X_B | X_C$ if and only if X_C d-separates X_A from X_B . (proof later in the quarter)

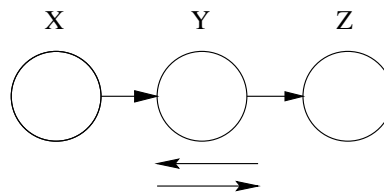
2.6 Bayes Ball Algorithm

This is an easy way to read off the conditional independence properties of a DGM i.e. to see if $X_A \perp\!\!\!\perp X_B | X_C$. Conditional independence is tested for as follows:

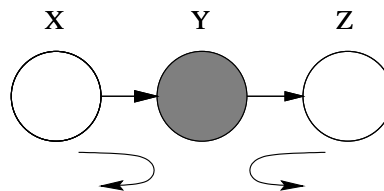
1. Shade all nodes of X_C
2. Place a ball in X_A and allow the ball to bounce on the graph according to certain rules concerning shaded nodes and the arrow direction (discussed below).
3. If it is possible for the ball to reach B starting in A , then $X_A \not\perp\!\!\!\perp X_B | X_C$. Otherwise, $X_A \perp\!\!\!\perp X_B | X_C$.

Bayes Ball Bouncing Rules

Case I:

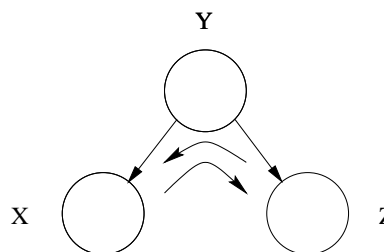


Ball goes through
Case I i.)

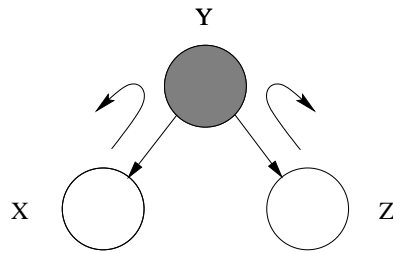


Ball is blocked
Case I ii.) $X \perp\!\!\!\perp Z | Y$

Case II:

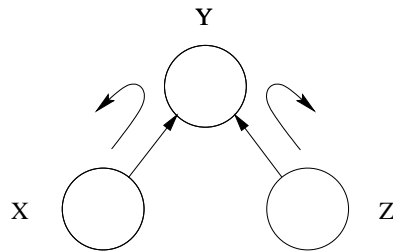


Case II i.)

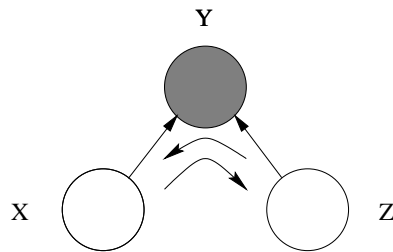


Case II ii.) $X \perp\!\!\!\perp Z|Y$

Case III:

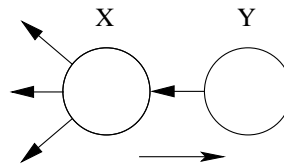


Case III i.) $X \perp\!\!\!\perp Y$



Case III ii.) $X \not\perp\!\!\!\perp Z|Y$

Boundary Conditions



Ball leaves the graph

Boundary i.)

