Review

- Conditional independence

\[ X \perp Y \mid Z \iff p(y, x \mid z) = p(x \mid z)p(y \mid z) \]

\[ \iff p(x \mid y, z) = p(x \mid z) \]

- Directed acyclic graph (DAG) \( G \): parents, children, descendents, non-descendents
- Bayes net: distribution is factorized. Each variable \( i \) “only depends on” it’s parents

\[ p(x) = \prod_{i=1}^{N} p(x_i \mid X_{pa(i)}) \]
Bayesian Networks

Conditional Independence and Factorization

Examples

D-Separation

Conditional Independence and Factorization

We assumed factorization in a Bayes net: \( p(x) = \prod_{i=1}^{N} p(x_i | x_{pa(i)}) \). What does this have to do with conditional independence?

Claim: for a probability distribution \( p(x) \)

\[
p(x) = \prod_{i=1}^{N} p(x_i | x_{pa(i)}) \iff X_i \perp X_{nd(i)} | X_{pa(i)} \text{ for all } i
\]

factorization \( \iff \) conditional independence

- RHS in words: conditionally independent of non-descendents given parents

Conditional Independence Implies Factorization

Assume \( X_i \perp X_{nd(i)} | X_{pa(i)} \) for all \( i \)

1. Number nodes according to topological ordering
   \( i \rightarrow j \Rightarrow i < j \)
   \( \Rightarrow \text{de}(i) \subseteq \{i+1, \ldots, N\} \)
   \( \Rightarrow \{i+1, \ldots, N\} \subseteq \text{nd}(i) \)

2. Chain rule
   \[
p(x_1, \ldots, x_n) = p(x_1) p(x_2|x_1) p(x_3|x_2) \ldots p(x_{N-1}|x_{N-2}, x_N)\]
   \[
   = \prod_{i=1}^{N} p(x_i | x_{pa(i)}, x_{\{i+1, \ldots, N\} \setminus \text{de}(i)})
   \]

3. Split into parents and other nd's

Review of Argument

0. Assume \( X_i \perp X_{nd(i)} | X_{pa(i)} \) for all \( i \)

1. Number nodes according to a topological ordering: \( i \rightarrow j \Rightarrow i < j \). Then we also have that \( \text{de}(i) \subseteq \{i+1, \ldots, n\} \), and, as a consequence all nodes in \( \{1, \ldots, i-1\} \) are non-descendents

2. Use the chain rule
   \[
p(x) = \prod_{i=1}^{N} p(x_i | x_{\{1, \ldots, i-1\}})
   \]

3. Split into parents and other non-descendents
   \[
p(x) = \prod_{i=1}^{N} p(x_i | x_{pa(i)} \setminus x_{\{1, \ldots, i-1\} \setminus \text{pa}(i)})
   \]

4. Simplify using conditional independence
   \[
p(x) = \prod_{i=1}^{N} p(x_i | x_{pa(i)})
   \]
Factorization Implies Conditional Independence

Assume \( p(x) = \prod_{i=1}^{N} p(x_i | x_{pa(i)}) \)

Fix \( i \) and write

\[
p(x) = \left( \prod_{j \in nd(i)} p(x_j | x_{pa(j)}) \right) \cdot p(x_i | x_{pa(i)}) \cdot p(x_{de(i)} | x_{nd(i)}, x_i)
\]

\[
= p(x_{nd(i)}) \cdot p(x_i | x_{pa(i)}) \cdot p(x_{de(i)} | x_{nd(i)}, x_i)
\]

Then we have

\[
p(x_i | x_{nd(i)}) = \frac{p(x_i, x_{nd(i)})}{p(x_{nd(i)})} = \frac{\sum_{x_{de(i)}} p(x_i, x_{nd(i)}, x_{de(i)})}{\sum_{x_{de(i)}} p(x_{nd(i)}, x_{de(i)})}
\]

\[
= \frac{\sum_{x_{de(i)}} p(x_{nd(i)}) \cdot p(x_i | x_{nd(i)}) \cdot p(x_{de(i)} | x_{nd(i)}, x_i)}{\sum_{x_{de(i)}} p(x_{nd(i)}) \cdot p(x_{de(i)} | x_{nd(i)}, x_i)}
\]

A Useful Fact

We discovered a useful fact twice in the preceding argument: marginalizing descendents in a Bayes net is very easy.

**Claim:** suppose \( A \) and \( B \) partition the nodes and there is no path from \( B \) to \( A \). Then

\[
p(x_A) = \sum_{x_B} p(x_A, x_B) = \prod_{i \in A} p(x_i | x_{pa(i)})
\]

Marginalizing non-descendents is not this simple.