Recall *algebraic* plausibility measures:

- Roughly speaking, they have operations $\oplus$ and $\otimes$ such that
  - $\text{Pl}(U \cup V) = \text{Pl}(U) \oplus \text{Pl}(V)$ if $U \cap V = \emptyset$
  - $\text{Pl}(U_1 \cap U_2 \mid U_3) = \text{Pl}(U_1 \mid U_2 \cap U_3) \otimes \text{Pl}(U_2 \mid U_3)$ if $U_2 \cap U_3 \in \mathcal{F}'$, $U_1, U_2, U_3 \in \mathcal{F}$.
Recall *algebraic* plausibility measures:

- Roughly speaking, they have operations $\oplus$ and $\otimes$ such that
  - $\text{Pl}(U \cup V) = \text{Pl}(U) \oplus \text{Pl}(V)$ if $U \cap V = \emptyset$
  - $\text{Pl}(U_1 \cap U_2 \mid U_3) = \text{Pl}(U_1 \mid U_2 \cap U_3) \otimes \text{Pl}(U_2 \mid U_3)$ if $U_2 \cap U_3 \in F', U_1, U_2, U_3 \in F$.

**Definition:** If $\text{Pl}$ is algebraic, then $U$ and $V$ do not interact given $V'$ if $\text{Pl}(U \cap V \mid V') = \text{Pl}(U \mid V') \otimes \text{Pl}(V \mid V')$ (if $V' \in F'$).
Recall \textit{algebraic} plausibility measures:

- Roughly speaking, they have operations $\oplus$ and $\otimes$ such that
  \begin{itemize}
  \item $\text{Pl}(U \cup V) = \text{Pl}(U) \oplus \text{Pl}(V)$ if $U \cap V = \emptyset$
  \item $\text{Pl}(U_1 \cap U_2 \mid U_3) = \text{Pl}(U_1 \mid U_2 \cap U_3) \otimes \text{Pl}(U_2 \mid U_3)$ if $U_2 \cap U_3 \in \mathcal{F}'$, $U_1, U_2, U_3 \in \mathcal{F}$.
  \end{itemize}

\textbf{Definition:} If $\text{Pl}$ is algebraic, then $U$ and $V$ do not interact given $V'$ if $\text{Pl}(U \cap V \mid V') = \text{Pl}(U \mid V') \otimes \text{Pl}(V \mid V')$ (if $V' \in \mathcal{F}'$).

\textbf{Lemma:} If $(W, \mathcal{F}, \mathcal{F}', \text{Pl})$ is an algebraic cps and either $U \cap V' \in \mathcal{F}'$ or $V \cap V' \in \mathcal{F}'$, then $I_{\text{Pl}}(U, V \mid V')$ implies $U$ and $V$ do not interact.
Recall *algebraic* plausibility measures:

- Roughly speaking, they have operations $\oplus$ and $\otimes$ such that
  - $\text{Pl}(U \cup V) = \text{Pl}(U) \oplus \text{Pl}(V)$ if $U \cap V = \emptyset$
  - $\text{Pl}(U_1 \cap U_2 \mid U_3) = \text{Pl}(U_1 \mid U_2 \cap U_3) \otimes \text{Pl}(U_2 \mid U_3)$ if $U_2 \cap U_3 \in \mathcal{F}'$, $U_1, U_2, U_3 \in \mathcal{F}$.

**Definition:** If $\text{Pl}$ is algebraic, then $U$ and $V$ *do not interact given* $V'$ if $\text{Pl}(U \cap V \mid V') = \text{Pl}(U \mid V') \otimes \text{Pl}(V \mid V')$ (if $V' \in \mathcal{F}'$).

**Lemma:** If $(W, \mathcal{F}, \mathcal{F}', \text{Pl})$ is an algebraic cps and either $U \cap V' \in \mathcal{F}'$ or $V \cap V' \in \mathcal{F}'$, then $I_{\text{Pl}}(U, V \mid V')$ implies $U$ and $V$ do not interact.

The converse is not necessarily true.

- There is an example in the book using possibility measures.
- We can an extra condition to get the converse

**Bottom line:** we can separate out the two notions of independence using algebraic plausibility measures.
Properties of independence for RVs

Recall:

CIRV1[µ]. If $I^\mu_{rv}(X, Y \mid Z)$, then $I^\mu_{rv}(Y, X \mid Z)$.

CIRV2[µ]. If $I^\mu_{rv}(X, Y \cup Y' \mid Z)$, then $I^\mu_{rv}(X, Y \mid Z)$.

CIRV3[µ]. If $I^\mu_{rv}(X, Y \cup Y' \mid Z)$, then $I^\mu_{rv}(X, Y \mid Y' \cup Z)$.

CIRV4[µ]. If $I^\mu_{rv}(X, Y \mid Z)$ and $I^\mu_{rv}(X, Y' \mid Y \cup Z)$, then $I^\mu_{rv}(X, Y \cup Y' \mid Z)$.

CIRV5[µ]. $I^\mu_{rv}(X, Z \mid Z)$. 

Theorem: These properties hold for all probability measures $\mu$.

More general theorem: If $Pl$ is an algebraic plausibility measure, then these properties continue to hold if we replace $I^\mu_{rv}$ with $I^\mu_{rv}(Pl)$. 
Properties of independence for RVs

Recall:

CIRV1[\mu]. \text{ If } I^{rv}_\mu(X, Y \mid Z), \text{ then } I^{rv}_\mu(Y, X \mid Z).

CIRV2[\mu]. \text{ If } I^{rv}_\mu(X, Y \cup Y' \mid Z), \text{ then } I^{rv}_\mu(X, Y \mid Z).

CIRV3[\mu]. \text{ If } I^{rv}_\mu(X, Y \cup Y' \mid Z), \text{ then } I^{rv}_\mu(X, Y \mid Y' \cup Z).

CIRV4[\mu]. \text{ If } I^{rv}_\mu(X, Y \mid Z) \text{ and } I^{rv}_\mu(X, Y' \mid Y \cup Z), \text{ then } I^{rv}_\mu(X, Y \cup Y' \mid Z).

CIRV5[\mu]. I^{rv}_\mu(X, Z \mid Z).

Theorem: These properties hold for all probability measures \mu.
Properties of independence for RVs

Recall:

**CIRV1[μ]**. If $I_{μ}^{rv}(X, Y \mid Z)$, then $I_{μ}^{rv}(Y, X \mid Z)$.

**CIRV2[μ]**. If $I_{μ}^{rv}(X, Y ∪ Y' \mid Z)$, then $I_{μ}^{rv}(X, Y \mid Z)$.

**CIRV3[μ]**. If $I_{μ}^{rv}(X, Y ∪ Y' \mid Z)$, then $I_{μ}^{rv}(X, Y \mid Y' ∪ Z)$.

**CIRV4[μ]**. If $I_{μ}^{rv}(X, Y \mid Z)$ and $I_{μ}^{rv}(X, Y' \mid Y ∪ Z)$, then $I_{μ}^{rv}(X, Y ∪ Y' \mid Z)$.

**CIRV5[μ]**. $I_{μ}^{rv}(X, Z \mid Z)$.

**Theorem**: These properties hold for all probability measures $μ$.

More general theorem:

**Theorem**: If $P_l$ is an algebraic plausibility measure, then these properties continue to hold if we replace $I_{μ}^{rv}$ with $I_{P_l}^{rv}$.
Qualitative Bayesian Networks

Recall: A *directed acyclic network* consists of a set of nodes and directed edges, where there are no cycles.

In a Bayesian network (BN), the nodes are labeled by random variables

We can think of the edges as representing causal influence
More definitions:

- The ancestors of $X$ in the graph are those random variables that have a potential influence on $X$.
  - $Y$ is an ancestor of $X$ in graph $G$ if there is a directed path from $Y$ to $X$ in $G$—i.e., a sequence $(Y_1, \ldots, Y_k)$ of nodes—such that $Y_1 = Y$, $Y_k = X$, and there is a directed edge from $Y_i$ to $Y_{i+1}$ for $i = 1, \ldots, k - 1$.

- The parents of $X$ in $G (\text{Par}_G(X))$ are those ancestors of $X$ directly connected to $X$.
  - $SH$ and $S$ are the parents of $C$, $PS$ is the parent of $S$

- The nondescendants of $X$ ($\text{NonDes}_G(X)$) are those nodes $Y$ such that $X$ is not the ancestor of $Y$. 
More definitions:

- **The ancestors** of $X$ in the graph are those random variables that have a potential influence on $X$.
  - $Y$ is an ancestor of $X$ in graph $G$ if there is a directed path from $Y$ to $X$ in $G$—i.e., a sequence $(Y_1, \ldots, Y_k)$ of nodes—such that $Y_1 = Y$, $Y_k = X$, and there is a directed edge from $Y_i$ to $Y_{i+1}$ for $i = 1, \ldots, k - 1$.

- **The parents** of $X$ in $G$ ($\text{Par}_G(X)$) are those ancestors of $X$ directly connected to $X$.
  - $SH$ and $S$ are the parents of $C$, $PS$ is the parent of $S$.

- **The nondescendants** of $X$ ($\text{NonDes}_G(X)$) are those nodes $Y$ such that $X$ is not the ancestor of $Y$.

**Key definition:** The Bayesian network $G$ (qualitatively) represents the probability measure $\mu$ if, for all nodes $X$ in $G$,

$$I^\mu (X, \text{NonDes}_G(X) | \text{Par}(X)).$$

- $X$ is independent of its nondescendants given its parents.
Why did we choose this weird definition of representation?

▶ It’s useful!
Why did we choose this weird definition of representation?
  ▶ It’s useful!

Suppose that a world is characterized by the value of the rvs $X_1, \ldots, X_n$, and we want to compute the probability of the world $(x_1, \ldots, x_n)$ without needing to store too many numbers.
  ▶ Knowing these conditional independencies let’s us do this
An apparent digression: the chain rule

Given arbitrary sets $U_1, \ldots, U_n$, it is immediate from the definition of conditional probability that

$$
\mu(U_1 \cap \ldots \cap U_n) = \mu(U_n \mid U_1 \cap \ldots \cap U_{n-1}) \times \mu(U_1 \cap \ldots \cap U_{n-1}).
$$
An apparent digression: the chain rule

Given arbitrary sets $U_1, \ldots, U_n$, it is immediate from the definition of conditional probability that

$$\mu(U_1 \cap \ldots \cap U_n) = \mu(U_n \mid U_1 \cap \ldots \cap U_{n-1}) \times \mu(U_1 \cap \ldots \cap U_{n-1}).$$

Applying this observation inductively gives the chain rule:

$$\mu(U_1 \cap \ldots \cap U_n) = \mu(U_n \mid U_1 \cap \ldots \cap U_{n-1}) \times \mu(U_{n-1} \mid U_1 \cap \ldots \cap U_{n-2}) \times \ldots \times \mu(U_2 \mid U_1) \times \mu(U_1).$$
An apparent digression: the chain rule

Given arbitrary sets $U_1, \ldots, U_n$, it is immediate from the definition of conditional probability that

$$\mu(U_1 \cap \ldots \cap U_n) = \mu(U_n \mid U_1 \cap \ldots \cap U_{n-1}) \times \mu(U_1 \cap \ldots \cap U_{n-1}).$$

Applying this observation inductively gives the chain rule:

$$\mu(U_1 \cap \ldots \cap U_n) = \mu(U_n \mid U_1 \cap \ldots \cap U_{n-1}) \times \mu(U_{n-1} \mid U_1 \cap \ldots \cap U_{n-2}) \times \ldots \times \mu(U_2 \mid U_1) \times \mu(U_1).$$

Now take $U_i$ be the event $X_i = x_i$.

- the set of all worlds where $X_i = x_i$

Plugging this into the chain rule gives:

$$\mu(x_1, \ldots, x_n) = \mu(X_1 = x_1 \cap \ldots \cap X_n = x_n)$$
$$= \mu(X_n = x_n \mid X_1 = x_1 \cap \ldots \cap X_{n-1} = x_{n-1}) \times \mu(X_{n-1} = x_{n-1} \mid X_1 = x_1 \cap \ldots \cap X_{n-2} = x_{n-2}) \times \ldots \times \mu(X_2 = x_2 \mid X_1 = x_1) \times \mu(X_1 = x_1).$$
To repeat, using the chain rule, we have:

\[
\mu(X_1 = x_1 \cap \ldots \cap X_n = x_n) = \mu(X_n = x_n \mid X_1 = x_1 \cap \ldots \cap X_{n-1} = x_{n-1}) \times \\
\mu(X_{n-1} = x_{n-1} \mid X_1 = x_1 \cap \ldots \cap X_{n-2} = x_{n-2}) \times \\
\ldots \times \mu(X_2 = x_2 \mid X_1 = x_1) \times \mu(X_1 = x_1).
\]
The Punch Line

To repeat, using the chain rule, we have:

\[
\mu(X_1 = x_1 \cap \ldots \cap X_n = x_n) \\
= \mu(X_n = x_n \mid X_1 = x_1 \cap \ldots \cap X_{n-1} = x_{n-1}) \times \\
\mu(X_{n-1} = x_{n-1} \mid X_1 = x_1 \cap \ldots \cap X_{n-2} = x_{n-2}) \times \\
\ldots \times \mu(X_2 = x_2 \mid X_1 = x_1) \times \mu(X_1 = x_1).
\]

Now suppose without loss of generality that \(\langle X_1, \ldots, X_n \rangle\) is a topological sort of (the nodes in) \(G\).

- if \(X_i\) is a parent of \(X_j\), then \(i < j\).

Thus, \(\{X_1, \ldots, X_{k-1}\} \subseteq \text{NonDes}_G(X_k)\), for \(k = 1, \ldots, n\)

- All the descendants of \(X_k\) must have subscripts \(> k\).

- Conclusion: all the nodes in \(\{X_1, \ldots, X_{k-1}\}\) are independent of \(X_k\) given \(\text{Par}_G(X_k)\).
The Punch Line

To repeat, using the chain rule, we have:

\[
\mu(X_1 = x_1 \cap \ldots \cap X_n = x_n)
\]
\[
= \mu(X_n = x_n \mid X_1 = x_1 \cap \ldots \cap X_{n-1} = x_{n-1}) \times 
\mu(X_{n-1} = x_{n-1} \mid X_1 = x_1 \cap \ldots \cap X_{n-2} = x_{n-2}) \times 
\ldots \times \mu(X_2 = x_2 \mid X_1 = x_1) \times \mu(X_1 = x_1).
\]

Now suppose without loss of generality that \(\langle X_1, \ldots, X_n \rangle\) is a topological sort of (the nodes in) \(G\).

▷ if \(X_i\) is a parent of \(X_j\), then \(i < j\).

Thus, \(\{X_1, \ldots, X_{k-1}\} \subseteq \text{NonDes}_G(X_k)\), for \(k = 1, \ldots, n\)

▷ All the descendants of \(X_k\) must have subscripts > \(k\).

▷ Conclusion: all the nodes in \(\{X_1, \ldots, X_{k-1}\}\) are independent of \(X_k\) given \(\text{Par}_G(X_k)\).

It follows that

\[
\mu(X_k = x_k \mid X_{k-1} = x_{k-1} \cap \ldots \cap X_1 = x_1)
\]
\[
= \mu(X_k = x_k \mid \cap_{x_i \in \text{Par}(X_k)} X_i = x_i).
\]
So we can greatly simplify our original equation:

\[
\mu(X_1 = x_1 \cap \ldots \cap X_n = x_n) = \\
= \mu(X_n = x_n \mid X_1 = x_1 \cap \ldots \cap X_{n-1} = x_{n-1}) \times \\
\mu(X_{n-1} = x_{n-1} \mid X_1 = x_1 \cap \ldots \cap X_{n-2} = x_{n-2}) \times \\
\ldots \times \mu(X_2 = x_2 \mid X_1 = x_1) \times \mu(X_1 = x_1).
\]

But if \( G \) represents \( \mu \), then

\[
\mu(X_k = x_k \mid X_{k-1} = x_{k-1} \cap \ldots \cap X_1 = x_1) = \\
= \mu(X_k = x_k \mid \cap X_i \in \text{Par}(X_k) X_i = x_i).
\]

So the equation above reduces to

\[
\mu(X_1 = x_1 \cap \ldots \cap X_n = x_n) = \\
= \mu(X_n = x_n \mid \cap X_i \in \text{Par}(X_n) X_i = x_i) \times \\
\mu(X_{n-1} = x_{n-1} \mid \cap X_i \in \text{Par}(X_{n-1}) X_i = x_i) \times \\
\ldots \times \mu(X_1 = x_1).
\]

Key point: If each variable \( X_i \) has relatively few parents, then to compute \( \mu(x_1, \ldots, x_n) \), we need relatively few numbers.
So we can greatly simplify our original equation:

\[ \mu(X_1 = x_1 \cap \ldots \cap X_n = x_n) = \mu(X_n = x_n \mid X_1 = x_1 \cap \ldots \cap X_{n-1} = x_{n-1}) \times \mu(X_{n-1} = x_{n-1} \mid X_1 = x_1 \cap \ldots \cap X_{n-2} = x_{n-2}) \times \ldots \times \mu(X_2 = x_2 \mid X_1 = x_1) \times \mu(X_1 = x_1). \]

But if \( G \) represents \( \mu \), then

\[ \mu(X_k = x_k \mid X_{k-1} = x_{k-1} \cap \ldots \cap X_1 = x_1) = \mu(X_k = x_k \mid \cap X_i \in \text{Par}(X_k) X_i = x_i). \]

So the equation above reduces to

\[ \mu(X_1 = x_1 \cap \ldots \cap X_n = x_n) = \mu(X_n = x_n \mid \cap X_i \in \text{Par}(X_n) X_i = x_i) \times \mu(X_{n-1} = x_{n-1} \mid \cap X_i \in \text{Par}(X_{n-1}) X_i = x_i) \times \ldots \times \mu(X_1 = x_1). \]

**Key point:** If each variable \( X_i \) has relatively few parents, then to compute \( \mu(x_1, \ldots, x_n) \), we need relatively few numbers.
Quantitative Bayesian Networks

A *quantitative Bayesian network* is a pair \((G, f)\) consisting of a qualitative Bayesian network \(G\) and a function \(f\) that associates with each node \(X\) in \(G\) a *conditional probability table (cpt)*. If \(\text{Par}_G(X) = Y\), then the cpt gives, for each possible setting \(x\) of \(X\) and \(y\) of \(Y\), a number \(f(X, x, Y, y)\).

\((G, f)\) *represents* \(\mu\) if

1. \(G\) qualitatively represents \(\mu\)
2. \(\mu(X = x \mid Y = y) = f(X, x, Y, y)\).
Quantitative Bayesian Networks

A quantitative Bayesian network is a pair \((G, f)\) consisting of a qualitative Bayesian network \(G\) and a function \(f\) that associates with each node \(X\) in \(G\) a conditional probability table (cpt). If \(\text{Par}_G(X) = Y\), then the cpt gives, for each possible setting \(x\) of \(X\) and \(y\) of \(Y\), a number \(f(X, x, Y, y)\).

\((G, f)\) represents \(\mu\) if

1. \(G\) qualitatively represents \(\mu\)

2. \(\mu(X = x \mid Y = y) = f(X, x, Y, y)\).

If \((G, f)\) quantitatively represents \(\mu\) then we can completely reconstruct \(\mu\) from \((G, f)\).

- Suppose that the world is described by \(N\) binary variables.
- This means that we are putting a probability distribution on \(2^N\) worlds.
- But if each rv has at most \(n\) parents, then each cpt requires at most \(2^{n+1}\) numbers
- At most \(N2^{n+1} \ll 2^N\) numbers needed altogether
**Example:** We get a quantitative BN for smoking by considering the qualitative BN:

![Graph](image)

together with the following cpts:

<table>
<thead>
<tr>
<th>$S$</th>
<th>$SH$</th>
<th>$C$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>.6</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>.4</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>.1</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>.01</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$PS$</th>
<th>$S$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>.4</td>
</tr>
<tr>
<td>0</td>
<td>.2</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$PS$</th>
<th>$SH$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>.8</td>
</tr>
<tr>
<td>0</td>
<td>.3</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$PS$</th>
</tr>
</thead>
<tbody>
<tr>
<td>.3</td>
</tr>
</tbody>
</table>
Constructing a Quantitative BN

**Proposition:** A quantitative BN $(G, f)$ represents a unique probability distribution $\mu$. 
Constructing a Quantitative BN

**Proposition:** A quantitative BN \((G, f)\) represents a unique probability distribution \(\mu\).

What about the converse? Given a probability distribution \(\mu\), can we find a quantitative BN that represents it?

- Yes! There are lots.

**Construction:**

- Given \(\mu\), let \(Y_1, \ldots, Y_n\) be any permutation of the random variables in \(\mathcal{X}\).
- For each \(k\), find a minimal subset of \(\{Y_1, \ldots, Y_{k-1}\}\), call it \(P_k\), such that \(I_{\mu}^{rv}(\{Y_1, \ldots, Y_{k-1}\}, Y_k \mid P_k)\).
  - There is a subset with this property, namely, \(\{Y_1, \ldots, Y_{k-1}\}\).
  - So there must be a minimal one
- Add edges from each of the nodes in \(P_k\) to \(Y_k\).
- Call the resulting graph \(G\).

**Theorem** \(G\) qualitatively represents \(\mu\).
Constructing a Quantitative BN

**Proposition:** A quantitative BN \((G, f)\) represents a unique probability distribution \(\mu\).

What about the converse? Given a probability distribution \(\mu\), can we find a quantitative BN that represents it?

▶ Yes! There are lots.

**Construction:**

▶ Given \(\mu\), let \(Y_1, \ldots, Y_n\) be any permutation of the random variables in \(\mathcal{X}\).

▶ For each \(k\), find a minimal subset of \(\{Y_1, \ldots, Y_{k-1}\}\), call it \(P_k\), such that \(I_{\mu}^{rv}(\{Y_1, \ldots, Y_{k-1}\}, Y_k \mid P_k)\).

▶ There is a subset with this property, namely, \(\{Y_1, \ldots, Y_{k-1}\}\).

▶ So there must be a minimal one

▶ Add edges from each of the nodes in \(P_k\) to \(Y_k\).

▶ Call the resulting graph \(G\).

**Theorem** \(G\) qualitatively represents \(\mu\).

▶ Now just add the “right” cpts
The Bayesian network constructed depends on the ordering of the edges.

- Different orderings may lead to different Bayesian networks.
  - The BN for smoking was constructed with the ordering $PS, S, SH, C$.
  - We could construct another one using the ordering $C, S, PS, SH$
    - It would have $C$ at the root
- Experience has shown that we get “better” BNs if we order the variables causally
  - If $X$ has a causal influence on $Y$, then $X$ precedes $Y$ in the order
    - This was the case with the original smoking network
  - “Better” typically means
    - fewer edges
    - easier to elicit the cpt from experts
The Bayesian network constructed depends on the ordering of the edges.

- Different orderings may lead to different Bayesian networks.
  - The BN for smoking was constructed with the ordering $PS, S, SH, C$.
  - We could construct another one using the ordering $C, S, PS, SH$
    - It would have $C$ at the root
  - Experience has shown that we get “better” BNs if we order the variables causally
    - If $X$ has a causal influence on $Y$, then $X$ precedes $Y$ in the order
      - This was the case with the original smoking network
  - “Better” typically means
    - fewer edges
    - easier to elicit the cpt from experts

This construction of BNs used only CIRV1-5

- Conclusion: it works without change for arbitrary algebraic plausibility measures
Independencies in BNs

If $G$ represents $\mu$, then an rv in $G$ is independent of its nondescendants conditional on its parents with respect to $\mu$.

- What other independencies hold?
- There is a criterion that lets us compute this.
d-separation

$X$ is $d$-separated ($d = \text{directed}$) from a node $Y$ by a set $Z$ of nodes in $G$, written $d-\text{sep}_G(X, Y \mid Z)$, if for every undirected path from $X$ to $Y$ there is a node $Z'$ on the path such that either

(a) $Z' \in Z$ and there is an arrow on the path leading in to $Z'$ and an arrow leading out from $Z'$;

(b) $Z' \in Z$ and has both path arrows leading out;

(c) $Z'$ has both path arrows leading in, and neither $Z'$ nor any of its descendants are in $Z$.

\[ X \quad Z \quad Y \]
d-separation

\(X\) is \textit{d-separated} (\(d = \text{directed}\)) from a node \(Y\) by a set \(Z\) of nodes in \(G\), written \(d\text{-sep}_G(X, Y \mid Z)\), if for every \textit{undirected path} from \(X\) to \(Y\) there is a node \(Z'\) on the path such that either

(a) \(Z' \in Z\) and there is an arrow on the path leading in to \(Z'\) and an arrow leading out from \(Z'\);

(b) \(Z' \in Z\) and has both path arrows leading out; or

\[ X \quad \quad Z \quad \quad Y \]
**d-separation**

$X$ is d-separated (\(d = \text{directed}\)) from a node $Y$ by a set $Z$ of nodes in $G$, written $d$-$\text{sep}_G(X, Y \mid Z)$, if for every undirected path from $X$ to $Y$ there is a node $Z'$ on the path such that either

(a) $Z' \in Z$ and there is an arrow on the path leading in to $Z'$ and an arrow leading out from $Z'$;

(b) $Z' \in Z$ and has both path arrows leading out;

(c) $Z'$ has both path arrows leading in, and neither $Z'$ nor any of its descendants are in $Z$.

\[
\begin{array}{c}
X \quad Z \quad Y \\
\end{array}
\]

\[
\begin{array}{c}
X \quad Z \quad Y \\
\end{array}
\]

\[
\begin{array}{c}
X \quad Z \quad Y \\
Y' \\
\end{array}
\]
X is \textit{d-separated} \((d = \text{directed})\) from a node \(Y\) by a set \(Z\) of nodes in \(G\), written \(d\text{-sep}_G(X, Y \mid Z)\), if for every \textit{undirected path} from \(X\) to \(Y\) there is a node \(Z'\) on the path such that either

(a) \(Z' \in Z\) and there is an arrow on the path leading in to \(Z'\) and an arrow leading out from \(Z'\);

(b) \(Z' \in Z\) and has both path arrows leading out; or

(c) \(Z'\) has both path arrows leading in, and neither \(Z'\) nor any of its descendants are in \(Z\).

Example:

- \(\{SH, S\}\) d-separates \(PS\) from \(C\).
- \(\{PS\}\) d-separates \(SH\) from \(S\).
- \(\{PS, C\}\) does \textit{not} d-separate \(SH\) from \(S\).
d-separation: some intuition

\( X \) is \textit{d-separated} (\( d = \text{directed} \)) from a node \( Y \) by a set \( Z \) of nodes in \( G \), written \( d\text{-sep}_G(X, Y \mid Z) \), if for every \textit{undirected path} from \( X \) to \( Y \) there is a node \( Z' \) on the path such that either

(a) \( Z' \in Z \) and there is an arrow on the path leading in to \( Z' \) and an arrow leading out from \( Z' \);

- pretty intuitive: conditioning on \( \{SH, S\} \) blocks all paths from \( PS \) to \( C \), so \( C \) is conditionally independent of \( PS \) given \( \{SH, S\} \).

(b) \( Z' \in Z \) and has both path arrows leading out;

- \( SH \) and \( S \) are not independent, because they have a common cause (\( PS \)), but conditioning on \( PS \) makes them independent

(c) \( Z' \) has both path arrows leading in, and neither \( Z' \) nor any of its descendants are in \( Z \);

- \( S \) and \( SH \) are independent conditional on \( PS \), but they would become dependent if we also conditioned on \( C \).
d-separation: some intuition

$X$ is $d$-separated ($d = \text{directed}$) from a node $Y$ by a set $Z$ of nodes in $G$, written $d$-$\text{sep}_G(X,Y \mid Z)$, if for every undirected path from $X$ to $Y$ there is a node $Z'$ on the path such that either

(a) $Z' \in Z$ and there is an arrow on the path leading in to $Z'$ and an arrow leading out from $Z'$;
   ▶ pretty intuitive: conditioning on $\{SH, S\}$ blocks all paths from $PS$ to $C$, so $C$ is conditionally independent of $PS$ given $\{SH, S\}$.

(b) $Z' \in Z$ and has both path arrows leading out; or
   ▶ $SH$ and $S$ are not independent, because they have a common cause ($PS$), but conditioning on $PS$ makes them independent

(c) $Z'$ has both path arrows leading in, and neither $Z'$ nor any of its descendants are in $Z$. 
d-separation: some intuition

$X$ is \textit{d-separated} \((d = \text{directed})\) from a node \(Y\) by a set \(Z\) of nodes in \(G\), written \(d\text{-sep}_G(X, Y \mid Z)\), if for every \textit{undirected path} from \(X\) to \(Y\) there is a node \(Z'\) on the path such that either

(a) \(Z' \in Z\) and there is an arrow on the path leading in to \(Z'\) and an arrow leading out from \(Z'\);

\begin{itemize}
  \item pretty intuitive: conditioning on \(\{SH, S\}\) blocks all paths from \(PS\) to \(C\), so \(C\) is conditionally independent of \(PS\) given \(\{SH, S\}\).
\end{itemize}

(b) \(Z' \in Z\) and has both path arrows leading out; or

\begin{itemize}
  \item \(SH\) and \(S\) are not independent, because they have a common cause \((PS)\), but conditioning on \(PS\) makes them independent
\end{itemize}

(c) \(Z'\) has both path arrows leading in, and neither \(Z'\) nor any of its descendants are in \(Z\).

\begin{itemize}
  \item \(S\) and \(SH\) are independent conditional on \(PS\), but they, would become \textit{dependent} if we also conditioned on \(C\)
\end{itemize}
D-separation completely characterizes conditional independence in Bayesian networks:

**Theorem:** If $X$ is d-separated from $Y$ by $Z$ in the Bayesian network $G$, then $I_{\mu}^{rv}(X, Y \mid Z)$ holds for all probability measures $\mu$ compatible with $G$. Conversely, if $X$ is not d-separated from $Y$ by $Z$, then there is a probability measure $\mu$ compatible with $G$ such that $I_{\mu}^{rv}(X, Y \mid Z)$ does not hold.
D-separation completely characterizes conditional independence in Bayesian networks:

**Theorem:** If \( X \) is d-separated from \( Y \) by \( Z \) in the Bayesian network \( G \), then \( I_{\mu}^r (X, Y \mid Z) \) holds for all probability measures \( \mu \) compatible with \( G \). Conversely, if \( X \) is *not* d-separated from \( Y \) by \( Z \), then there is a probability measure \( \mu \) compatible with \( G \) such that \( I_{\mu}^r (X, Y \mid Z) \) does not hold.

- The proof of the first half of the theorem requires only CIRV1-5, so holds for all algebraic plausibility measures.
- For the second half, we need some extra conditions.

**Bottom line:** the technology of Bayesian networks can be applied quite widely!