Recall *algebraic* plausibility measures:

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

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

Recall *algebraic* plausibility measures:

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

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

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

Recall *algebraic* plausibility measures:

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

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

Lemma: If $(W, \mathcal{F}, \mathcal{F}', \mathrm{Pl})$ is an algebraic cps and either $U \cap V' \in \mathcal{F}'$ or $V \cap V' \in \mathcal{F}'$, then $I_{\mathrm{Pl}}(U, V \mid V')$ 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:

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

$\mathrm{CIRV2}[\mu]$. If $I_\mu^{rv}(\mathbf{X},\mathbf{Y} \cup \mathbf{Y}' \mid \mathbf{Z})$, then $I_\mu^{rv}(\mathbf{X},\mathbf{Y} \mid \mathbf{Z})$.

$\mathrm{CIRV3}[\mu]$. If $I_\mu^{rv}(\mathbf{X},\mathbf{Y} \cup \mathbf{Y}' \mid \mathbf{Z})$, then $I_\mu^{rv}(\mathbf{X},\mathbf{Y} \mid \mathbf{Y}' \cup \mathbf{Z})$.

$\mathrm{CIRV4}[\mu]$. If $I_\mu^{rv}(\mathbf{X},\mathbf{Y} \mid \mathbf{Z})$ and $I_\mu^{rv}(\mathbf{X},\mathbf{Y}' \mid \mathbf{Y} \cup \mathbf{Z})$, then $I_\mu^{rv}(\mathbf{X},\mathbf{Y} \cup \mathbf{Y}' \mid \mathbf{Z})$.

$\mathrm{CIRV5}[\mu]$. $I_\mu^{rv}(\mathbf{X},\mathbf{Z} \mid \mathbf{Z})$.

Properties of independence for RVs

Recall:

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

$\mathrm{CIRV2}[\mu]$. If $I_\mu^{rv}(\mathbf{X},\mathbf{Y} \cup \mathbf{Y}' \mid \mathbf{Z})$, then $I_\mu^{rv}(\mathbf{X},\mathbf{Y} \mid \mathbf{Z})$.

$\mathrm{CIRV3}[\mu]$. If $I_\mu^{rv}(\mathbf{X},\mathbf{Y} \cup \mathbf{Y}' \mid \mathbf{Z})$, then $I_\mu^{rv}(\mathbf{X},\mathbf{Y} \mid \mathbf{Y}' \cup \mathbf{Z})$.

$\mathrm{CIRV4}[\mu]$. If $I_\mu^{rv}(\mathbf{X},\mathbf{Y} \mid \mathbf{Z})$ and $I_\mu^{rv}(\mathbf{X},\mathbf{Y}' \mid \mathbf{Y} \cup \mathbf{Z})$, then $I_\mu^{rv}(\mathbf{X},\mathbf{Y} \cup \mathbf{Y}' \mid \mathbf{Z})$.

$\mathrm{CIRV5}[\mu]$. $I_\mu^{rv}(\mathbf{X},\mathbf{Z} \mid \mathbf{Z})$.

Theorem: These properties hold for all probability measures μ.

Properties of independence for RVs

Recall:

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

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

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

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

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

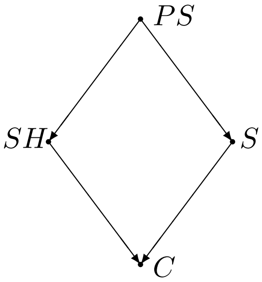
Theorem: These properties hold for all probability measures μ.

More general theorem:

Theorem: If Pl is an algebraic plausibility measure, then these properties continue to hold if we replace I_μ^{rv} with I_{Pl}^{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 $(\mathrm{Par}_G(X))$ 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 $(\mathrm{NonDes}_G(X))$ 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 $(\mathrm{Par}_G(X))$ 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 $(\mathrm{NonDes}_G(X))$ are those nodes Y such that X is not the ancestor of Y.

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

$$I_\mu^{rv}(X, \mathrm{NonDes}_G(X) \mid \mathrm{Par}(X)).$$

- X is independent of its nondescendants given its parents

Why did we choose this weird definition of represntation?

- It's useful!

Why did we choose this weird definition of represntation?

- 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:

$$\begin{aligned} & \mu(U_1 \cap \ldots \cap U_n) \\ = \; & \mu(U_n \mid U_1 \cap \ldots \cap U_{n-1}) \times \\ & \quad \mu(U_{n-1} \mid U_1 \cap \ldots \cap U_{n-2}) \times \ldots \times \mu(U_2 \mid U_1) \times \mu(U_1). \end{aligned}$$

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:

$$\begin{aligned} & \mu(U_1 \cap \ldots \cap U_n) \\ = \; & \mu(U_n \mid U_1 \cap \ldots \cap U_{n-1}) \times \\ & \quad \mu(U_{n-1} \mid U_1 \cap \ldots \cap U_{n-2}) \times \ldots \times \mu(U_2 \mid U_1) \times \mu(U_1). \end{aligned}$$

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:

$$\begin{aligned} & \mu(x_1, \ldots, x_n) \\ = \; & \mu(X_1 = x_1 \cap \ldots \cap X_n = x_n) \\ = \; & \mu(X_n = x_n \mid X_1 = x_1 \cap \ldots \cap X_{n-1} = x_{n-1}) \times \\ & \quad \mu(X_{n-1} = x_{n-1} \mid X_1 = x_1 \cap \ldots \cap X_{n-2} = x_{n-2}) \times \\ & \quad \ldots \times \mu(X_2 = x_2 \mid X_1 = x_1) \times \mu(X_1 = x_1). \end{aligned}$$

The Punch Line

To repeat, using the chain rule, we have:

$$
\begin{aligned}
& \mu(X_1 = x_1 \cap \ldots \cap X_n = x_n) \\
= \; & \mu(X_n = x_n \mid X_1 = x_1 \cap \ldots \cap X_{n-1} = x_{n-1}) \times \\
& \quad \mu(X_{n-1} = x_{n-1} \mid X_1 = x_1 \cap \ldots \cap X_{n-2} = x_{n-2}) \times \\
& \quad \ldots \times \mu(X_2 = x_2 \mid X_1 = x_1) \times \mu(X_1 = x_1).
\end{aligned}
$$

The Punch Line

To repeat, using the chain rule, we have:

$$
\begin{aligned}
& \mu(X_1 = x_1 \cap \ldots \cap X_n = x_n) \\
= \;& \mu(X_n = x_n \mid X_1 = x_1 \cap \ldots \cap X_{n-1} = x_{n-1}) \times \\
& \quad \mu(X_{n-1} = x_{n-1} \mid X_1 = x_1 \cap \ldots \cap X_{n-2} = x_{n-2}) \times \\
& \quad \ldots \times \mu(X_2 = x_2 \mid X_1 = x_1) \times \mu(X_1 = x_1).
\end{aligned}
$$

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, $\{X_1, \ldots, X_{k-1}\} \subseteq \mathrm{NonDes}_G(X_k)$, 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 $\mathrm{Par}_G(X_k)$.

The Punch Line

To repeat, using the chain rule, we have:

$$\begin{array}{rl} & \mu(X_1 = x_1 \cap \ldots \cap X_n = x_n) \\ = & \mu(X_n = x_n \mid X_1 = x_1 \cap \ldots \cap X_{n-1} = x_{n-1}) \times \\ & \quad \mu(X_{n-1} = x_{n-1} \mid X_1 = x_1 \cap \ldots \cap X_{n-2} = x_{n-2}) \times \\ & \quad \ldots \times \mu(X_2 = x_2 \mid X_1 = x_1) \times \mu(X_1 = x_1). \end{array}$$

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, $\{X_1, \ldots, X_{k-1}\} \subseteq \mathrm{NonDes}_G(X_k)$, 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 $\mathrm{Par}_G(X_k)$.

It follows that

$$\begin{array}{rl} & \mu(X_k = x_k \mid X_{k-1} = x_{k-1} \cap \ldots \cap X_1 = x_1) \\ = & \mu(X_k = x_k \mid \cap_{X_i \in \mathrm{Par}(X_k)} X_i = x_i). \end{array}$$

So we can greatly simplify our original equation:

$$\begin{aligned} & \mu(X_1 = x_1 \cap \ldots \cap X_n = x_n) \\ = \; & \mu(X_n = x_n \mid X_1 = x_1 \cap \ldots \cap X_{n-1} = x_{n-1}) \times \\ & \quad \mu(X_{n-1} = x_{n-1} \mid X_1 = x_1 \cap \ldots \cap X_{n-2} = x_{n-2}) \times \\ & \quad \ldots \times \mu(X_2 = x_2 \mid X_1 = x_1) \times \mu(X_1 = x_1). \end{aligned}$$

But if G represents μ, then

$$\begin{aligned} & \mu(X_k = x_k \mid X_{k-1} = x_{k-1} \cap \ldots \cap X_1 = x_1) \\ = \; & \mu(X_k = x_k \mid \cap_{X_i \in \mathrm{Par}(X_k)} X_i = x_i). \end{aligned}$$

So the equation above reduces to

$$\begin{aligned} & \mu(X_1 = x_1 \cap \ldots \cap X_n = x_n) \\ = \; & \mu(X_n = x_n \mid \cap_{X_i \in \mathrm{Par}(X_n)} X_i = x_i) \times \\ & \quad \mu(X_{n-1} = x_{n-1} \mid \cap_{X_i \in \mathrm{Par}(X_{n-1})} X_i = x_i) \times \\ & \qquad \cdots \times \mu(X_1 = x_1). \end{aligned}$$

So we can greatly simplify our original equation:

$$\begin{aligned} & \mu(X_1 = x_1 \cap \ldots \cap X_n = x_n) \\ = \; & \mu(X_n = x_n \mid X_1 = x_1 \cap \ldots \cap X_{n-1} = x_{n-1}) \times \\ & \quad \mu(X_{n-1} = x_{n-1} \mid X_1 = x_1 \cap \ldots \cap X_{n-2} = x_{n-2}) \times \\ & \quad \ldots \times \mu(X_2 = x_2 \mid X_1 = x_1) \times \mu(X_1 = x_1). \end{aligned}$$

But if G represents μ, then

$$\begin{aligned} & \mu(X_k = x_k \mid X_{k-1} = x_{k-1} \cap \ldots \cap X_1 = x_1) \\ = \; & \mu(X_k = x_k \mid \cap_{X_i \in \mathrm{Par}(X_k)} X_i = x_i). \end{aligned}$$

So the equation above reduces to

$$\begin{aligned} & \mu(X_1 = x_1 \cap \ldots \cap X_n = x_n) \\ = \; & \mu(X_n = x_n \mid \cap_{X_i \in \mathrm{Par}(X_n)} X_i = x_i) \times \\ & \quad \mu(X_{n-1} = x_{n-1} \mid \cap_{X_i \in \mathrm{Par}(X_{n-1})} X_i = x_i) \times \\ & \quad\quad \cdots \times \mu(X_1 = x_1). \end{aligned}$$

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 $\mathrm{Par}_G(X) = \mathbf{Y}$, then the cpt gives, for each possible setting x of X and $\mathbf{y}$ of $\mathbf{Y}$, a number $f(X, x, \mathbf{Y}, \mathbf{y})$.

(G, f) *represents* μ if

1. G qualitatively represents μ
2. $\mu(X = x \mid \mathbf{Y} = \mathbf{y}) = f(X, x, \mathbf{Y}, \mathbf{y})$.

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 $\mathrm{Par}_G(X) = \mathbf{Y}$, then the cpt gives, for each possible setting x of X and $\mathbf{y}$ of $\mathbf{Y}$, a number $f(X, x, \mathbf{Y}, \mathbf{y})$.

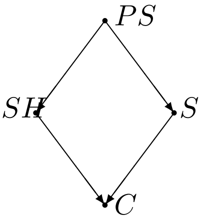
(G, f) *represents* μ if

1. G qualitatively represents μ
2. $\mu(X = x \mid \mathbf{Y} = \mathbf{y}) = f(X, x, \mathbf{Y}, \mathbf{y})$.

If (G, f) quantitatively represents μ then we can completely reconstruct μ 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:



together with the following cpts:

S	SH	C
1	1	.6
1	0	.4
0	1	.1
0	0	.01

PS	S
1	.4
0	.2

PS	SH
1	.8
0	.3

PS
.3

Constructing a Quantative BN

Proposition: A quantititive BN (G, f) represents a unique probability distribution μ.

Constructing a Quantative BN

Proposition: A quantititive BN (G, f) represents a unique probability distribution μ.

What about the converse? Given a probabilty distribution μ, can we find a quantitative BN that represents it?

- Yes! There are lots.

Construction:

- Given μ, 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 $\mathbf{P}_k$, such that $I_\mu^{rv}(\{Y_1, \ldots, Y_{k-1}\}, Y_k \mid \mathbf{P}_k)$.
  - 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 $\mathbf{P}_k$ to Y_k.
- Call the resulting graph G.

Theorem G qualitatively represents μ.

Constructing a Quantative BN

Proposition: A quantititive BN (G, f) represents a unique probability distribution μ.

What about the converse? Given a probabilty distribution μ, can we find a quantitative BN that represents it?

- Yes! There are lots.

Construction:

- Given μ, 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 $\mathbf{P}_k$, such that $I_\mu^{rv}(\{Y_1, \ldots, Y_{k-1}\}, Y_k \mid \mathbf{P}_k)$.
  - 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 $\mathbf{P}_k$ to Y_k.
- Call the resulting graph G.

Theorem G qualitatively represents μ.

- Now just add the "right" cpts

The Bayesian network construted 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 construted 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 μ, then an rv in G is independent of its nondescendants conditional on its parents with respect to μ.

- What other independencies hold?
- There is a critrion that lets us compute this.

d-separation

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

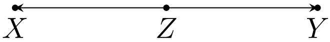
(a) $Z' \in \mathbf{Z}$ and there is an arrow on the path leading in to Z' and an arrow leading out from Z';

d-separation

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

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

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

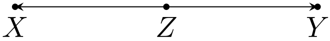


d-separation

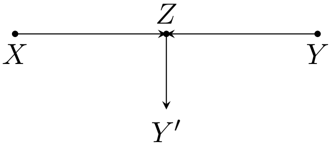
X is *d-separated* (*d* = *directed*) from a node Y by a set $\mathbf{Z}$ of nodes in G, written *d-sep*$_G(X, Y \mid \mathbf{Z})$, if for every *undirected path* from X to Y there is a node Z' on the path such that either

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

(b) $Z' \in \mathbf{Z}$ 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 $\mathbf{Z}$.



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

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

(b) $Z' \in \mathbf{Z}$ 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 $\mathbf{Z}$.

Example:

- $\{SH, S\}$ d-separates PS from C.
- $\{PS\}$ d-separates SH from S.
- $\{PS, C\}$ does *not* d-separate SH from S.

d-separation: some intuition

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

(a) $Z' \in \mathbf{Z}$ 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 $\{SH, S\}$.

d-separation: some intuition

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

(a) $Z' \in \mathbf{Z}$ 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) $Z' \in \mathbf{Z}$ 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 $\mathbf{Z}$.

d-separation: some intuition

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

(a) $Z' \in \mathbf{Z}$ 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) $Z' \in \mathbf{Z}$ 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 $\mathbf{Z}$.
   - S and SH are indepedent conditional on PS, but they, would become *dependent* if we also conditioned on C

D-separation completely characterizes conditional independence in Bayesian networks:

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

D-separation completely characterizes conditional independence in Bayesian networks:

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