Graphical Models For Complex Health Data (P8124)

Daniel Malinsky

Columbia University
dsm2128@cumc.columbia.edu

Directed Graphical Models

Directed Acyclic Graphs

A directed acyclic graph (DAG)is a graph $\mathcal{G} = (V, E)$ is such that E contains only directed edges ($\rightarrow$) and has no sequence of directed edges from X_i to X_i ($\forall i$).

Definitions:

$$
\begin{aligned}
\text{Pa}(X_i, \mathcal{G}) &\equiv \{X_j : X_j \rightarrow X_i \text{ in } \mathcal{G}\} && \text{(parents of } X_i\text{)} \\
\text{Ch}(X_i, \mathcal{G}) &\equiv \{X_j : X_j \leftarrow X_i \text{ in } \mathcal{G}\} && \text{(children of } X_i\text{)} \\
\text{An}(X_i, \mathcal{G}) &\equiv \{X_j : X_j \rightarrow \ldots \rightarrow X_i \text{ in } \mathcal{G}\} && \text{(ancestors of } X_i\text{)} \\
\text{De}(X_i, \mathcal{G}) &\equiv \{X_j : X_j \leftarrow \ldots \leftarrow X_i \text{ in } \mathcal{G}\} && \text{(descendants of } X_i\text{)}
\end{aligned}
$$

Factorization Property

A distribution $p(x)$ satisfies the *factorization property* wrt DAG $\mathcal{G}$ if

$$p(x) = \prod_{i=1}^{p} p(x_i | \operatorname{Pa}(X_i, \mathcal{G}))$$

where $\operatorname{Pa}(X_i, \mathcal{G})$ are the parents of X_i in $\mathcal{G}$.

Definition: Bayesian network model

A pair $(\mathcal{G}, \mathcal{P})$ where $\mathcal{G}$ is a DAG and $\mathcal{P}$ is a set of distributions that factorize wrt $\mathcal{G}$.

Chain rule vs. factorization wrt DAG

By the chain rule of probability, any distribution

$$
\begin{aligned}
p(x) &= p(x_1, ..., x_p) \\
&= p(x_p | x_{p-1}, ..., x_1) p(x_{p-1}, ..., x_1) \\
&= p(x_p | x_{p-1}, ..., x_1) p(x_{p-1} | x_{p-2}, ..., x_1) p(x_{p-2}, ..., x_1) \\
&= \prod_{i=1}^{p} p(x_i | x_{i-1}, ..., x_1)
\end{aligned}
$$

Chain rule vs. factorization wrt DAG

By the chain rule of probability, any distribution

$$\begin{aligned} p(x) &= p(x_1, ..., x_p) \\ &= p(x_p | x_{p-1}, ..., x_1) p(x_{p-1}, ..., x_1) \\ &= p(x_p | x_{p-1}, ..., x_1) p(x_{p-1} | x_{p-2}, ..., x_1) p(x_{p-2}, ..., x_1) \\ &= \prod_{i=1}^{p} p(x_i | x_{i-1}, ..., x_1) \end{aligned}$$

Compare this to the DAG factorization:
$p(x) = \prod_{i=1}^{p} p(x_i | \mathrm{Pa}(X_i, \mathcal{G}))$
What does the factorization property "get you"?

Chain rule vs. factorization wrt DAG

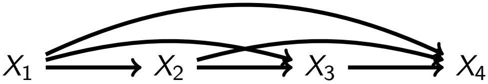


Figure: A complete DAG

$$
\begin{aligned}
p(x) &= \prod_{i=1}^{p} p(x_i \,|\, \mathrm{Pa}(X_i, \mathcal{G})) \\
&= p(x_1)p(x_2|x_1)p(x_3|x_2, x_1)p(x_4|x_3, x_2, x_1)
\end{aligned}
$$

Chain rule vs. factorization wrt DAG

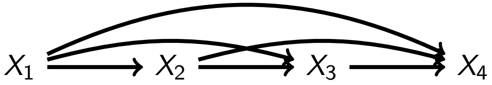


Figure: A complete DAG

$$
\begin{aligned}
p(x) &= \prod_{i=1}^{p} p(x_i \,|\, \mathrm{Pa}(X_i, \mathcal{G})) \\
&= p(x_1)p(x_2|x_1)p(x_3|x_2, x_1)p(x_4|x_3, x_2, x_1)
\end{aligned}
$$

For a complete DAG the factorization property doesn't say anything! Just equivalent to chain rule.

$\Rightarrow$ "A complete DAG imposes no restrictions on the data distribution."

Chain rule vs. factorization wrt DAG

$$X_1 \longrightarrow X_2 \longrightarrow X_3 \longrightarrow X_4$$

Figure: A chain

$$\begin{aligned} p(x) &= \prod_{i=1}^{p} p(x_i \mid \mathrm{Pa}(X_i, \mathcal{G})) \\ &= p(x_1)p(x_2|x_1)p(x_3|x_2)p(x_4|x_3) \end{aligned}$$

Chain rule vs. factorization wrt DAG

$$X_1 \longrightarrow X_2 \longrightarrow X_3 \longrightarrow X_4$$

Figure: A chain

$$\begin{aligned} p(x) &= \prod_{i=1}^{p} p(x_i \mid \mathrm{Pa}(X_i, \mathcal{G})) \\ &= p(x_1)p(x_2|x_1)p(x_3|x_2)p(x_4|x_3) \end{aligned}$$

A DAG with missing edges encodes conditional independencies.

Which conditional independencies?

Compare:

$$\begin{aligned} p(x) &= p(x_1)p(x_2|x_1)p(x_3|x_2,x_1)p(x_4|x_3,x_2,x_1) \\ &= p(x_1)p(x_2|x_1)p(x_3|x_2)p(x_4|x_3) \end{aligned}$$

$p(x_4|x_3,x_2,x_1) = p(x_4|x_3)$ or $X_4 \perp\!\!\!\perp X_2, X_1|X_3$
and
$p(x_3|x_2,x_1) = p(x_3|x_2)$ or $X_3 \perp\!\!\!\perp X_1|X_2$

Which conditional independencies?

Compare:

$$
\begin{aligned}
p(x) &= p(x_1)p(x_2|x_1)p(x_3|x_2,x_1)p(x_4|x_3,x_2,x_1) \\
&= p(x_1)p(x_2|x_1)p(x_3|x_2)p(x_4|x_3)
\end{aligned}
$$

$p(x_4|x_3,x_2,x_1) = p(x_4|x_3)$ or $X_4 \perp\!\!\!\perp X_2, X_1 | X_3$
and
$p(x_3|x_2,x_1) = p(x_3|x_2)$ or $X_3 \perp\!\!\!\perp X_1 | X_2$

something like... $X_i \perp\!\!\!\perp$ "ancestors" $|$ parents ??

Local Markov property

Define:

$$\text{Nd}(X_i, \mathcal{G}) \equiv \{X_j : X_j \notin \text{De}(X_i, \mathcal{G})\} \qquad \text{(non-descendants of } X_i\text{)}$$

Local Markov property

Define:

$$\mathrm{Nd}(X_i, \mathcal{G}) \equiv \{X_j : X_j \notin \mathrm{De}(X_i, \mathcal{G})\} \qquad \text{(non-descendants of } X_i\text{)}$$

A distribution $p(x)$ satisfies the *local Markov property* wrt DAG $\mathcal{G}$ if

$X_i \perp\!\!\!\perp \mathrm{Nd}^*(X_i, \mathcal{G}) \mid \mathrm{Pa}(X_i, \mathcal{G})$

where $\mathrm{Nd}^*(X_i, \mathcal{G}) \equiv \mathrm{Nd}(X_i, \mathcal{G}) \setminus \mathrm{Pa}(X_i, \mathcal{G})$

Local Markov property

$$\begin{tikzcd}
X_1 \arrow[r] & X_2 \arrow[r] & X_3 \arrow[r] & X_4 \\
 & & X_5 \arrow[ur] &
\end{tikzcd}$$

$$X_4 \perp\!\!\!\perp X_2, X_1 | X_3, X_5$$
$$X_3 \perp\!\!\!\perp X_1, X_5 | X_2$$
$$\dots$$

Global Markov property

Let A, B, C be disjoint subsets of X. A distribution $p(x)$ satisfies the *global Markov property* wrt DAG $\mathcal{G}$ if

$A \perp_d B|C \implies A \perp\!\!\!\perp B|C$.

Global Markov property

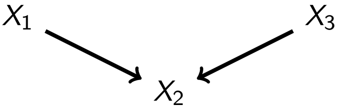
Let A, B, C be disjoint subsets of X. A distribution $p(x)$ satisfies the *global Markov property* wrt DAG $\mathcal{G}$ if

$A \perp_d B | C \implies A \perp\!\!\!\perp B | C.$

"A graphical criterion $\implies$ conditional independence in the distribution"

Colliders

Given a path π in a graph $\mathcal{G}$, a non-endpoint vertex X_j on π is called a collider if the two edges incident to X_j are both into X_j, i.e., have arrowheads at X_j.[1]



[1] A v-structure is a triple $\langle X_i, X_j, X_k \rangle$ such that X_j is a collider and X_i and X_k are not adjacent. A collider which is part of a v-structure (i.e., a collider with non-adjacent parents) is also called an unshielded collider.

d-separation

A path π in $\mathcal{G}$ between distinct vertices X_i and X_k is called a *d-connecting path* conditional on vertex set C ($C \subseteq X \setminus \{X_i, X_k\}$) if:
a) every collider on π is in C or is in $\text{An}(C, \mathcal{G})$ and
b) every non-collider on π is not in C.

X_i and X_k are *d-separated* conditional on C if there is no d-connecting path conditional on C between X_i and X_k.

$A \perp_d B | C$ if $X_i \perp_d X_k | C$ for all $X_i \in A$ and $X_k \in B$.

d-separation (equivalent def)

A vertex X_j is **active** on a path relative to C just in case either
i) X_j is a collider, and X_j or any of its descendents is in C, or
ii) X_j is a noncollider and is not in C.

d-separation (equivalent def)

A vertex X_j is **active** on a path relative to C just in case either
i) X_j is a collider, and X_j or any of its descendents is in C, or
ii) X_j is a noncollider and is not in C.

A path π is **active** relative to C just in case every vertex on π is active relative to C.

d-separation (equivalent def)

A vertex X_j is **active** on a path relative to C just in case either
i) X_j is a collider, and X_j or any of its descendents is in C, or
ii) X_j is a noncollider and is not in C.

A path π is **active** relative to C just in case every vertex on π is active relative to C.

X_i and X_k are d-separated given C just in case there is no path between X_i and X_k that is **active** relative to C.

$\Rightarrow$ "Active can be thought of as carrying association"

d-separation examples

[on the board]

Equivalence of Markov properties in BNs

Theorem. Let $\mathcal{G}$ be a DAG. For any probability distribution which has a density wrt product measure, the factorization, local Markov, and Global Markov properities (wrt $\mathcal{G}$) are equivalent:

Factorization $\iff$ Local Markov $\iff$ Global Markov

Prove: global $\implies$ local

Need to show that $X_i \perp_d \text{Nd}(X_i) \setminus \text{Pa}(X_i) | \text{Pa}(X_i)$. Then global property $\implies X_i \perp\!\!\!\perp \text{Nd}(X_i) \setminus \text{Pa}(X_i) | \text{Pa}(X_i)$ (local).

Any path from X_i to $\text{Nd}(X_i)$ must go through either $\text{Ch}(X_i)$ or $\text{Pa}(X_i)$. If the path goes through $\text{Ch}(X_i)$ then there must be a collider on that path, but it is not in the conditioning set $\text{Pa}(X_i)$ (nor an ancestor of $\text{Pa}(X_i)$ by acyclicity), so that path must be not active. If instead the path goes through $\text{Pa}(X_i)$, the parents would be non-colliders on the path, and since they are in the conditionining set, they are also not active. So all paths are not active, therefore the paths cannot be d-connecting.

The reverse direction (global $\impliedby$ local) is trickier: proof by induction on graph size.

I-map

Let $\mathcal{P}$ be a set of distributions (model) over X. We define $\mathcal{I}(\mathcal{P})$ to be the set of independence assertions of the form $(A \perp\!\!\!\perp B|C)$ that hold in $\mathcal{P}$. $(A, B, C$ are disjoint subsets of X.)

I-map

Let $\mathcal{P}$ be a set of distributions (model) over X. We define $\mathcal{I}(\mathcal{P})$ to be the set of independence assertions of the form $(A \perp\!\!\!\perp B | C)$ that hold in $\mathcal{P}$. (A, B, C are disjoint subsets of X.)

Let $\mathcal{I}(\mathcal{G})$ be the set of independencies implied by the (local or global) Markov property for $\mathcal{G}$. We say that $\mathcal{G}$ is an I-map of $\mathcal{I}(\mathcal{P})$ if $\mathcal{I}(\mathcal{G}) \subseteq \mathcal{I}(\mathcal{P})$.

I-map

Let $\mathcal{P}$ be a set of distributions (model) over X. We define $\mathcal{I}(\mathcal{P})$ to be the set of independence assertions of the form $(A \perp\!\!\!\perp B|C)$ that hold in $\mathcal{P}$. (A, B, C are disjoint subsets of X.)

Let $\mathcal{I}(\mathcal{G})$ be the set of independencies implied by the (local or global) Markov property for $\mathcal{G}$. We say that $\mathcal{G}$ is an I-map of $\mathcal{I}(\mathcal{P})$ if $\mathcal{I}(\mathcal{G}) \subseteq \mathcal{I}(\mathcal{P})$.

Note: $\mathcal{P}$ may have independencies that are not reflected in $\mathcal{I}(\mathcal{G})$!

Example: deterministic relationships among variables

Consider:

$$X_1 \longrightarrow X_2 \longrightarrow X_3$$

where $X_3 = 2 \times X_2$.

Example: deterministic relationships among variables

Consider:

$$X_1 \longrightarrow X_2 \longrightarrow X_3$$

where $X_3 = 2 \times X_2$.

$X_3 \perp\!\!\!\perp X_1 | X_2$ (by Markov property) but also $X_2 \perp\!\!\!\perp X_1 | X_3$ (by determinism).

Example: exact "cancellation" or "balancing"

Consider:

$$\begin{tikzcd} X_1 \arrow[r] \arrow[rr, bend left] & X_2 \arrow[r] & X_3 \end{tikzcd}$$

where

$$\begin{aligned} X_1 &= \epsilon_1 \\ X_2 &= \alpha X_1 + \epsilon_2 \\ X_3 &= \beta X_2 - \alpha\beta X_1 + \epsilon_3 \\ \epsilon_1, \epsilon_2, \epsilon_3 &\sim N(0,1), \quad \alpha, \beta > 0 \end{aligned}$$

Example: exact "cancellation" or "balancing"

Consider:

$$\begin{tikzcd} X_1 \arrow[r] \arrow[rr, bend left] & X_2 \arrow[r] & X_3 \end{tikzcd}$$

where

$$\begin{aligned} X_1 &= \epsilon_1 \\ X_2 &= \alpha X_1 + \epsilon_2 \\ X_3 &= \beta X_2 - \alpha\beta X_1 + \epsilon_3 \\ \epsilon_1, \epsilon_2, \epsilon_3 &\sim N(0,1), \quad \alpha, \beta > 0 \end{aligned}$$

No independencies follow from Markov property, but $X_3 \perp\!\!\!\perp X_1$ (by exact cancellation).

The Faithfulness Assumption

It is common to rule out "extra" or "non-graphical" conditional independencies, *by assumption*. This restricts the allowed set of distributions, and may not always be appropriate.

A distribution $p(x)$ satisfies the *faithfulness assumption* wrt DAG $\mathcal{G}$ if

$A \perp\!\!\!\perp B|C \implies A \perp_d B|C.$

The Faithfulness Assumption

It is common to rule out "extra" or "non-graphical" conditional independencies, *by assumption*. This restricts the allowed set of distributions, and may not always be appropriate.

A distribution $p(x)$ satisfies the *faithfulness assumption* wrt DAG $\mathcal{G}$ if

$A \perp\!\!\!\perp B|C \implies A \perp_d B|C$.

In conjunction with the global Markov property this means we're assuming $A \perp\!\!\!\perp B|C \iff A \perp_d B|C$.

The Faithfulness Assumption

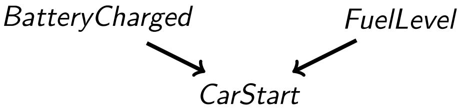
... is not always an appropriate assumption to make, but greatly simplifies things ("all conditional independence information is in the graph") and plays an important role in structure learning (as we will discuss later).

The Faithfulness Assumption

... is not always an appropriate assumption to make, but greatly simplifies things ("all conditional independence information is in the graph") and plays an important role in structure learning (as we will discuss later).

Also note that if a distribution satisfies the Markov and faithfulness assumptions wrt $\mathcal{G}$ we sometimes say that $\mathcal{G}$ is a *perfect map* for that distribution.

Colliders, Chains, Forks



Colliders seem special. Why?

Colliders, Chains, Forks

$$\begin{tikzcd} \mathit{BatteryCharged} \arrow[dr] & & \mathit{FuelLevel} \arrow[dl] \\ & \mathit{CarStart} & \end{tikzcd}$$

Colliders seem special. Why?

By d-separation: $\mathit{BatteryCharged} \perp\!\!\!\perp \mathit{FuelLevel}$.

Colliders, Chains, Forks

$$\begin{tikzcd} \mathit{BatteryCharged} \arrow[dr] & & \mathit{FuelLevel} \arrow[dl] \\ & \mathit{CarStart} & \end{tikzcd}$$

Colliders seem special. Why?

By d-separation: $\mathit{BatteryCharged} \perp\!\!\!\perp \mathit{FuelLevel}$.
By faithfulness assumption: $\mathit{BatteryCharged} \not\perp\!\!\!\perp \mathit{FuelLevel} | \mathit{CarStart}$.

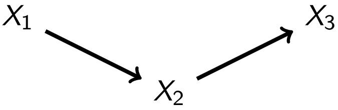
Conditioning on *CarStart* makes *FuelLevel* informative about *BatteryCharged*.

Colliders, Chains, Forks

$$X_1 \rightarrow X_2 \rightarrow X_3$$

Very different from "chains."

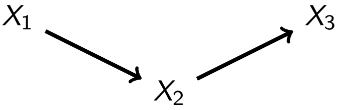
Colliders, Chains, Forks



Very different from "chains."

By d-separation: $X_1 \not\perp\!\!\!\perp X_3$.

Colliders, Chains, Forks



Very different from “chains.”

By d-separation: $X_1 \not\perp\!\!\!\perp X_3$.
By d-separation: $X_1 \perp\!\!\!\perp X_3 | X_2$.

Conditioning on X_2 makes X_1 non-informative about X_3.

Colliders, Chains, Forks

$$X_1 \leftarrow X_2 \rightarrow X_3$$

And "forks."

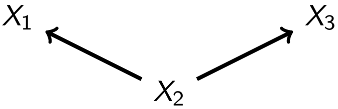
Colliders, Chains, Forks

$$X_1 \leftarrow X_2 \rightarrow X_3$$

And "forks."

By d-separation: $X_1 \not\perp\!\!\!\perp X_3$.

Colliders, Chains, Forks



And "forks."

By d-separation: $X_1 \not\perp\!\!\!\perp X_3$.
By d-separation: $X_1 \perp\!\!\!\perp X_3 | X_2$.

Conditioning on X_2 makes X_1 non-informative about X_3.

Markov equivalence

$$
\begin{array}{cccc}
X_1 \to X_2 \to X_3 & X_1 \leftarrow X_2 \leftarrow X_3 & X_1 \leftarrow X_2 \to X_3 & X_1 \to X_2 \leftarrow X_3 \\
\Downarrow & \Downarrow & \Downarrow & \Downarrow \\
X_1 \perp\!\!\!\perp X_3 | X_2 & X_1 \perp\!\!\!\perp X_3 | X_2 & X_1 \perp\!\!\!\perp X_3 | X_2 & X_1 \perp\!\!\!\perp X_3 \\
 & & & X_1 \not\!\perp\!\!\!\perp X_3 | X_2 \\
 & & & \text{w/ faithfulness}
\end{array}
$$

Markov equivalence

Definition. Two DAGs $\mathcal{G}_1$ and $\mathcal{G}_2$ are called Markov equivalent if $\mathcal{I}(\mathcal{G}_1) = \mathcal{I}(\mathcal{G}_2)$.

Theorem. Two DAGs $\mathcal{G}_1$ and $\mathcal{G}_2$ are Markov equivalent iff they share the same adjacencies and unshielded colliders.

Markov equivalence

a)	b)	c)
$X_i \rightarrow X_j \rightarrow X_k$		
$X_i \leftarrow X_j \leftarrow X_k$		
$X_i \leftarrow X_j \rightarrow X_k$	$X_i - X_j - X_k$	$X_i \rightarrow X_j \leftarrow X_k$

Figure: a) Three Markov equivalent DAG models. b) The CPDAG representation of (a). c) A DAG that is not Markov equivalent to the graphs in (a).

Implications of Markov equivalence

Markov equivalence is *almost* like "observational equivalence" – the data cannot distinguish between Markov equivalent graphs (unless we use more than conditional independence information).

So, using only conditional independence information, **we cannot learn the "correct" structure within an equivalence class**. Equivalence classes are the natural "units" for structure learning. (More on this later!)

Markov blankets

$\text{Mb}(X_i, \mathcal{G})$ is called the Markov blanket of X_i. It is a set of vertices that "screens off" all other vertices in $\mathcal{G}$, i.e., $X_i \perp\!\!\!\perp X \setminus \{\text{Mb}(X_i, \mathcal{G}), X_i\} \,|\, \text{Mb}(X_i, \mathcal{G})$.

In a DAG, $\text{Mb}(X_i, \mathcal{G}) \equiv \text{Pa}(X_i) \cup \text{Ch}(X_i) \cup \text{Pa}(\text{Ch}(X_i))$.

Markov blankets

$\mathrm{Mb}(X_i, \mathcal{G})$ is called the Markov blanket of X_i. It is a set of vertices that "screens off" all other vertices in $\mathcal{G}$, i.e.,
$X_i \perp\!\!\!\perp X \setminus \{\mathrm{Mb}(X_i, \mathcal{G}), X_i\} \,|\, \mathrm{Mb}(X_i, \mathcal{G})$.

In a DAG, $\mathrm{Mb}(X_i, \mathcal{G}) \equiv \mathrm{Pa}(X_i) \cup \mathrm{Ch}(X_i) \cup \mathrm{Pa}(\mathrm{Ch}(X_i))$.

If you want only to predict X_i, $\mathrm{Mb}(X_i, \mathcal{G})$ is sufficient.
[example on board]

Question: can every distribution be described by a BN?

Question: can every distribution be described by a BN?

Yes, in a boring sense: just use the complete DAG.

Question: can every distribution be described by a BN?

Yes, in a boring sense: just use the complete DAG.

The complete DAG is an I-map of any model, since $\mathcal{I}(\mathcal{G})$ is empty.

Question: can every distribution be described by a BN *faithfully*?

Question: can every distribution be described by a BN *faithfully*?

No!

Question: can every distribution be described by a BN *faithfully*?

No! Even leaving issues such as determinism and exact cancellation aside, there are sets of conditional independence facts that correspond to no DAG over the *observed* variables.

Consider:
$X_1 \perp\!\!\!\perp X_3 | (X_2, X_4)$ and $X_2 \perp\!\!\!\perp X_4 | (X_1, X_3)$
(and all vars are pairwise dependent)

Question: can every distribution be described by a BN *faithfully*?

No! Even leaving issues such as determinism and exact cancellation aside, there are sets of conditional independence facts that correspond to no DAG over the *observed* variables.

Consider:
$X_1 \perp\!\!\!\perp X_3 | (X_2, X_4)$ and $X_2 \perp\!\!\!\perp X_4 | (X_1, X_3)$
(and all vars are pairwise dependent)
$\Rightarrow$ there is no faithful BN to describe these independencies (may use a MRF)

Question: can every distribution be described by a BN *faithfully*?

Consider:

$X_1 \not\perp\!\!\!\perp X_2$ and $X_2 \not\perp\!\!\!\perp X_3$ and $X_3 \not\perp\!\!\!\perp X_4$

$X_1 \perp\!\!\!\perp X_4$ and $X_1 \perp\!\!\!\perp X_3$ and $X_2 \perp\!\!\!\perp X_4$

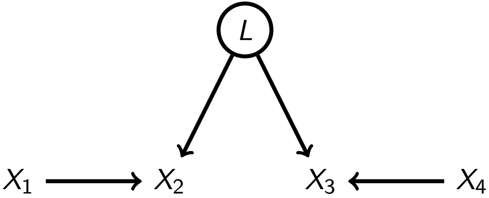
$X_1 \not\perp\!\!\!\perp X_3 | X_2$

$X_2 \not\perp\!\!\!\perp X_4 | X_3$

Question: can every distribution be described by a BN *faithfully*?

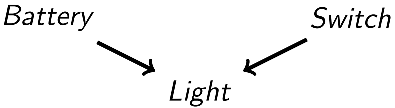
Consider:

$X_1 \not\perp\!\!\!\perp X_2$ and $X_2 \not\perp\!\!\!\perp X_3$ and $X_3 \not\perp\!\!\!\perp X_4$
$X_1 \perp\!\!\!\perp X_4$ and $X_1 \perp\!\!\!\perp X_3$ and $X_2 \perp\!\!\!\perp X_4$
$X_1 \not\perp\!\!\!\perp X_3 | X_2$
$X_2 \not\perp\!\!\!\perp X_4 | X_3$



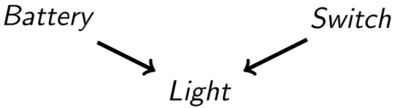
$\Rightarrow$ Might have to posit *unobserved* variables to explain observed independencies

Context-specific independence



Consider a battery-powered light bulb, with a switch. Charge in the battery will cause the bulb to light up provided the switch is on, but not otherwise. The dependence of *Light* and *Battery* arises entirely through the condition *Switch* = 'on'. When *Switch* = 'off', *Light* and *Battery* are independent.

Context-specific independence



Consider a battery-powered light bulb, with a switch. Charge in the battery will cause the bulb to light up provided the switch is on, but not otherwise. The dependence of *Light* and *Battery* arises entirely through the condition *Switch* = 'on'. When *Switch* = 'off', *Light* and *Battery* are independent.

These independencies are correctly represented by the simple DAG, but it is not fully informative. The independence is "context-specific," i.e., only appears in the context of a certain variable setting. Representing this system with a single DAG isn't wrong, but there are alternative, richer representations that may be more useful. Likewise for "interactions."

Context-specific independence

$$\begin{array}{ccc} & \textit{Switch} = \text{``}\textit{on}\text{''} & \\ \textit{Battery} & \longrightarrow & \textit{Light} \end{array} \qquad \begin{array}{ccc} & \textit{Switch} = \text{``}\textit{off}\text{''} & \\ \textit{Battery} & & \textit{Light} \end{array}$$

We could represent this situation using a *set* of Bayesian Networks, a.k.a. a multinet. One BN for every “context” setting.

Special structure: Hidden Markov Models (HMMs)

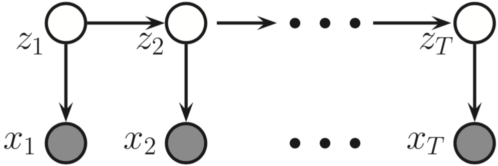


Figure 10.4 A first-order HMM.

(from Murphy (2012))

Special structure: Dynamic Bayesian Networks (DBNs)

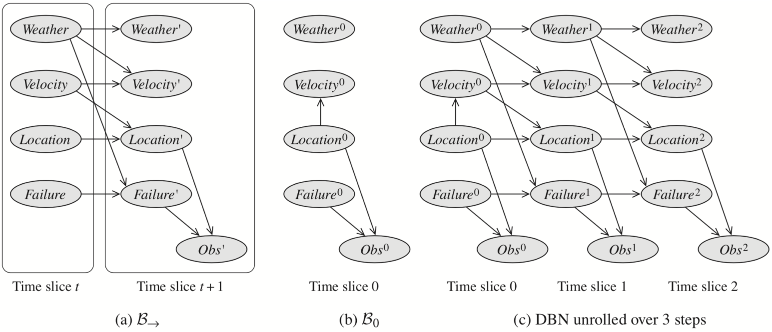


Figure 6.1 A highly simplified DBN for monitoring a vehicle: (a) the 2-TBN; (b) the time 0 network; (c) resulting unrolled DBN over three time slices.