Independence, Conditional Independence and d-Separation

Sources: 1. Elements of Causal Inference, (Peters, Janzing, Scholkopf), Chap 6. (aka "text")

2. Notes by M. Meila @ UW (Stat 535)

3. Intro to Bayesian Networks, N. L. Zhang (slides/notes), HKUST COMP 538.

4. Handbook of Graphical Models, Maathuis, Drton, Lauritzen, Wainwright, (mostly chap 1, Milan Studený)

Setting: $(X_1, X_2, \cdots, X_d)$ a collection of random variables, with joint distribution $P_{\underline{X}}(\cdot)$, i.e.,

$$P_{X_1,\cdots X_d}(x_1, \cdots, x_d) = P(X_1 = x_1, \cdots, X_d = x_d)$$

(we deal with discrete below for notation simplicity; the analog with pdfs holds for cont. r.v.s)

Henceforth, $A_i \equiv \{X_i = x_i\}$, $i = 1, 2, \cdots d$ (abusing notation here, because strictly we need to have the notation $A_i(x_i)$, as x_i is a parameter.)

Then, $P_{X_1,\cdots X_d}(x_1, \cdots, x_d) = P\left(\bigcap_{i=1}^{d} A_i\right)$.

Chain Rule: $P\left(\bigcap_{i=1}^{d} A_i\right) = P(A_1)\,P(A_2|A_1)\cdots\cdots P(A_d|A_1, \cdots A_{d-1})$

Notation: $A_i \perp\!\!\!\perp A_j$ ↪ "independent of"

$$A_i \perp\!\!\!\perp A_j \mid A_k \quad : \quad P(A_i \cap A_j \mid A_k)$$

indep. of ↙ ↓ conditioned on

$$= P(A_i \mid A_k)\, P(A_j \mid A_k)$$

Lemma 1: $A_i \perp\!\!\!\perp A_j \mid A_k \iff \exists\ \phi_{ik}, \phi_{jk}$ s.t.

$$P(A_i \cap A_j \cap A_k) = \phi_{ik}(x_i, x_k) \cdot \phi_{jk}(x_j, x_k)$$

Pf: $(\Rightarrow)$

$$P(A_i \cap A_j \cap A_k) \overset{\uparrow}{=} \frac{P(A_i \cap A_k)}{\sqrt{P(A_k)}} \cdot \frac{P(A_j \cap A_k)}{\sqrt{P(A_k)}}$$

(use $A_i \perp\!\!\!\perp A_j \mid A_k$)

$$\boxed{\begin{aligned} &= P(A_i \mid A_k)\, P(A_j \mid A_k)\, P(A_k) \\ &= \frac{P(A_i \cap A_k)}{P(A_k)} \frac{P(A_j \cap A_k)}{P(A_k)} \cdot P(A_k) \end{aligned}}$$

$(\Leftarrow)$

$$P(A_i \cap A_k) = \sum_{x_j} \phi_{ik}(x_i, x_k)\, \phi_{jk}(x_j, x_k)$$

$$= \phi_{ik}(x_i, x_k) \cdot \gamma_1(x_k) \quad — ①$$

$$P(A_j \cap A_k) = \phi_{jk}(x_j, x_k)\, \gamma_2(x_k). \quad — ②$$

Observe $A_i \perp\!\!\!\perp A_j \mid A_k \iff \dfrac{P(A_i \cap A_k)\, P(A_j \cap A_k)}{P(A_k)} = P(A_i \cap A_j \cap A_k)$

From ①, ②:

$$\frac{P(A_i \cap A_k)\, P(A_j \cap A_k)}{P(A_k)} = \frac{\phi_{ik}(x_i, x_k) \cdot \phi_{jk}(x_j, x_k)\, \cancel{\gamma_1(x_k)}\, \cancel{\gamma_2(x_k)}}{\cancel{\gamma_1(x_k)}\, \cancel{\gamma_2(x_k)}}$$

$$= P(A_i \cap A_j \cap A_k).$$

$$P(A_k) = \sum_j P(A_j \cap A_k)$$

$$= \underbrace{\sum_{x_j} \phi_{jk}(x_j, x_k) \cdot \gamma_2(x_k)}_{= \gamma_1(x_k)}$$

$\blacksquare$

Directed Acyclic Graph (DAG) and CI

$\mathcal{G} = (V, E)$, $|V| = d$,

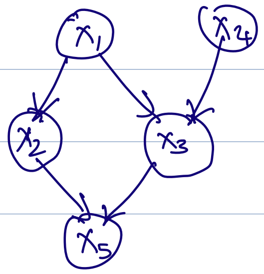
nodes in the DAG represents each one of the r.vs $\{X_1, X_2, \cdots, X_d\}$; edges encode dependency defined as follows:

For node j (i.e., r.v. X_j), PA_j are its parents on the directed graph. Then, a DAG encodes the following factorization of the joint distribution

$$P\left(\bigcap_{j=1}^{d} A_j\right) = \prod_{j=1}^{d} P(A_j \mid \underset{\hookrightarrow \{X_j = x_j\}}{PA_j})$$

Notation: $p(x_1, \dots, x_d)$ is the pmf (pdf) of the r.vs.

Example:

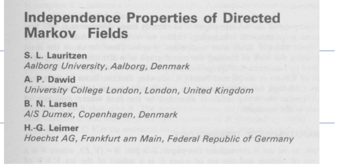


$$p(x_1, x_2, x_3, x_4, x_5) =$$

$$p(x_1)\, p(x_2 | x_1)\, p(x_3 | x_1, x_4)\, p(x_4)\, p(x_5 | x_2, x_3)$$

High Level Goal #1:

Additional discussion:



Independence Properties of Directed Markov Fields

S. L. Lauritzen
Aalborg University, Aalborg, Denmark

A. P. Dawid
University College London, London, United Kingdom

B. N. Larsen
A/S Dumex, Copenhagen, Denmark

H.-G. Leimer
Hoechst AG, Frankfurt am Main, Federal Republic of Germany

Given the DAG, can we "read off" conditional independence relations among the random variables?

---

e.g. In figure above, $p(x_2, x_3 | x_1, x_4)$

$$\overset{?}{=} p(x_2 | x_1, x_4)\, p(x_3 | x_1, x_4)$$

i.e., is the following statement true: $X_2 \perp\!\!\!\perp X_3 \,|\, X_1, X_4$?

Other examples: $X_2 \perp\!\!\!\perp X_3$?

$X_2 \perp\!\!\!\perp X_3 \,|\, X_5$?

"Problem" : Conditional independence is a (non-intuitive) algebraic property

Roadmap:

ⓐ Define a different independence operation $\perp\!\!\!\perp_G$, called "d-separation" that is defined by structural conditions on the DAG G.

↳ graphical property.

ⓑ (Global Markov Property)

$$X_i \perp\!\!\!\perp_G X_j \mid X_k \implies X_i \perp\!\!\!\perp X_j \mid X_k$$

Show that for the models we will use for causal inference, Global Markov Property holds.

ⓒ Discuss faithfulness: $X_i \perp\!\!\!\perp X_j \mid X_k \implies X_i \perp\!\!\!\perp_G X_j \mid X_k$

---

d-Separation

(Ref: d-Separation without tears, J. Pearl)

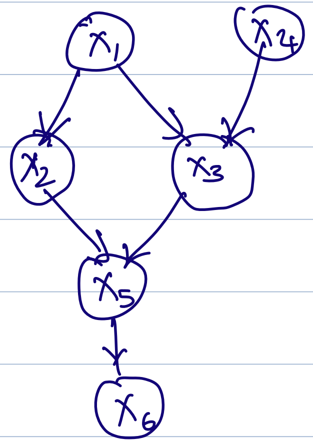
Language: Path : undirected sequence of edges from

$$X_i - X_k - X_\ell - X_j$$

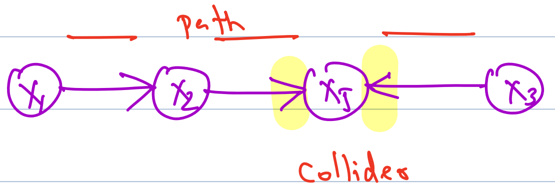
(with no repeated nodes)

(for a path)

Collider: A node in the path s.t. both edge arrows are incident on it.



ex:

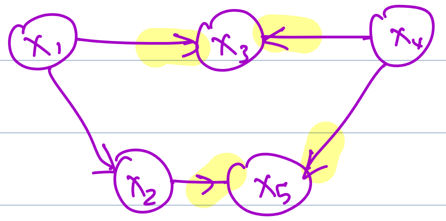


Blocked path: Path has a collider

Rules for d-separation:

ⓐ $X_i \perp\!\!\!\perp_G X_j$ if every path between X_i and X_j is blocked.

ex:

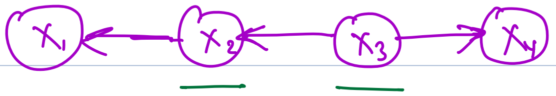


only 2 paths between X_1 and X_4

Note: All statements below are for each specific path. A node may be a collider along one path but not a collider along another.

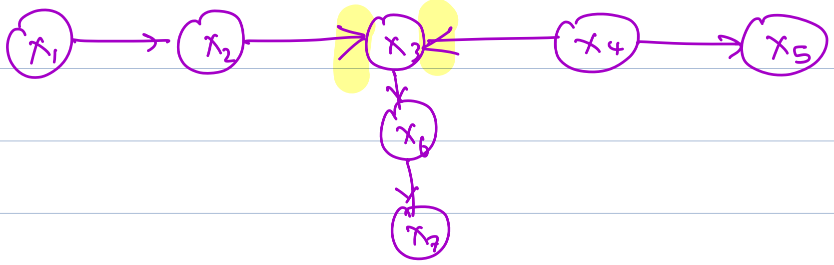
ⓑ Conditioning: $X_i \perp\!\!\!\perp_G X_j \mid (X_k, X_l \cdots)$.

<u>b. 1</u>: Conditioning on a non-collider node blocks the path



Conditioning on any or all of $\{X_2, X_3\}$ blocks the path between X_1 and X_4.

<u>b. 2</u>: Conditioning on a collider node or any of its descendents unblocks the path.



X_3 is a collider, and blocks the path between X_1 and X_5, <u>if</u> there is no conditioning

BUT Conditioning on any of $\{X_3, X_6, X_7\}$ unblocks the path between X_1 and X_5.

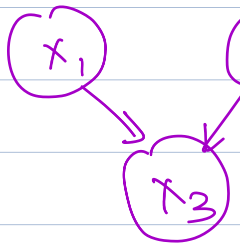
i.e. $X_1 \perp\!\!\!\perp_G X_5$ but

$$X_1 \not\perp\!\!\!\perp_G X_5 \mid \{X_3, X_6, X_7\}$$

any subset of above

Remark: Conditioning on a collider leaks information across its two parents.

example:



$X_1 = \text{Bernoulli}(0.5)$
$X_2 = \text{Bernoulli}(0.5)$

$X_3 = (X_1 \oplus X_2)$.

exclusive OR

Then, even if X_1 and X_2 are independent, conditioning on X_3 makes (X_1, X_2) dependent.

Summary: $X_i \perp\!\!\!\perp_G X_j \mid \{X_{k_1}, \dots X_{k_\ell}\}$ if there are no unblocked paths between X_i and X_j.

Remark: For other closely related graphical ways to reason about CI, see: Lauritzen et. al., Independence properties of directed Markov random fields, Networks 1990.

→ (Theorem 2 below)

Theorem: $X_i \perp\!\!\!\perp_G X_j \mid \{X_{k_1}, \dots X_{k_\ell}\}$

$$\Longrightarrow X_i \perp\!\!\!\perp X_j \mid \{X_{k_1}, \dots, X_{k_\ell}\}.$$

(This is called the Global Markov Property)

Proof: (Closely follows N. Zhang's notes)

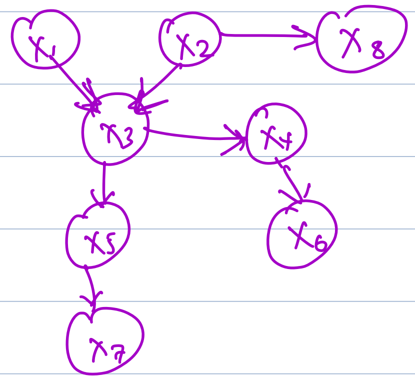
$\{X_1, \dots X_d\}$ nodes on the DAG.

$\mathcal{X} \subseteq \{X_1, \dots, X_d\}$.

Defns: X_i a leaf node if X_i has no children.

ancestor$(\mathcal{X}) = \mathcal{X} \cup \{X_k : \exists$ directed path from X_k to some node in $\mathcal{X}\}$.

$\mathcal{X}$ is ancestral if ancestor$(\mathcal{X}) = \mathcal{X}$.



ancestor $\{X_5\}$ example

$= \{X_5, X_3, X_1, X_2\}$

X_7 is a leaf node

Lemma 2: $\mathcal{G}$ a DAG, and X_k a leaf node. Then, let $\mathcal{G}' = \mathcal{G} \setminus X_k$, i.e., the DAG with the leaf node removed. Let $\mathcal{Y} = \{X_1, \dots X_d\} \setminus X_k$.

Then, $P_{\mathcal{G}}(\mathcal{Y}) = P_{\mathcal{G}'}(\mathcal{Y})$

where $P_{\mathcal{G}}(\mathcal{Y})$ is the joint dist. of the variables in $\mathcal{Y}$ derived from marginizing the joint dist on $\mathcal{G}$.

More precisely,

$$P_{\mathcal{G}'}(\mathcal{Y}) = \underset{\text{marginalize out } X_k}{\sum_{x_k}} \underbrace{\left(\prod_{i=1}^{d} p(x_i \mid PA_i) \right)}_{\text{joint dist. of } \{X_1, \dots X_d\} \text{ under DAG } \mathcal{G}. \to P_{\mathcal{G}}(\mathcal{Y})}$$

$$P_{\mathcal{G}'}(\mathcal{Y}) = \prod_{\substack{i=1 \\ d \neq k}}^{d} p(x_i \mid PA_i).$$

<u>Pf:</u> Immediate, as $\sum_{x_k} p(x_k \mid PA_k) = 1.$

∎

Lemma 3: Suppose $\mathcal{X}$ is ancestral. Let $\mathcal{G}'$ be the DAG with all nodes outside $\mathcal{X}$ removed.

Then, $P_G(\mathcal{X}) = P_{G'}(\mathcal{X})$.

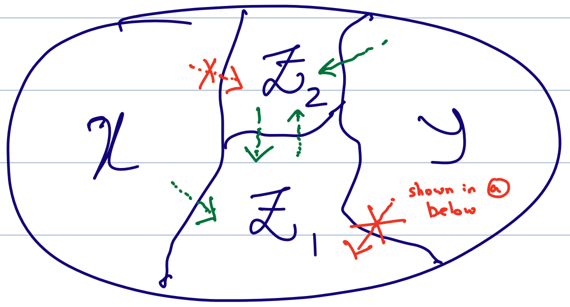
Proof: Find a leaf outside of $\mathcal{X}$; remove. Recursively do this, and we will remove all nodes outside of $\mathcal{X}$. (follows from ancestral property).
Now apply Lemma 2 repeatedly through each removal step above. $\boxed{\mathcal{E}}$

Theorem 1: Suppose $\mathcal{X}, \mathcal{Y}, \mathcal{Z}$ partitions $\{x_1, \cdots, x_d\}$.

Then,

$$\mathcal{X} \perp\!\!\!\perp_G \mathcal{Y} \mid \mathcal{Z}$$

$$\Rightarrow \mathcal{X} \perp\!\!\!\perp \mathcal{Y} \mid \mathcal{Z}$$



$\mathcal{Z} = \mathcal{Z}_1 \cup \mathcal{Z}_2$
with $\mathcal{Z}_1 \cap \mathcal{Z}_2 = \emptyset$.

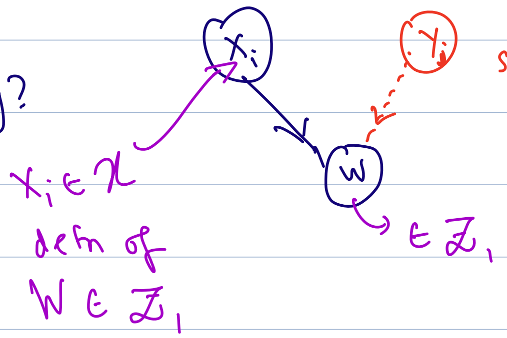
Proof: $\mathcal{Z}_1 \subseteq \mathcal{Z}$ s.t. $\exists$ parent$(W) \in \mathcal{X}$, $\forall$ $W \in \mathcal{Z}_1$

↳ note: this means that atleast one of the parents of $W \in \mathcal{X}$; we will show below that none of the parents of $W \in \mathcal{Y}$.

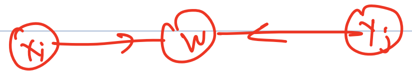
$\mathcal{Z}_2 = \mathcal{Z} \setminus \mathcal{Z}_1$

ⓐ For any $W \in \mathcal{X} \cup \mathcal{Z}_1$, observe that:
parent$(W) \subseteq \mathcal{X} \cup \mathcal{Z}$ (specifically, parent$(W) \notin \mathcal{Y}$.)

Why?



suppose $Y_j \in \mathcal{Y}$ was a parent of W.
Then, W is a collider for



$\Rightarrow$ $\mathcal{Z}$ not a d-separator

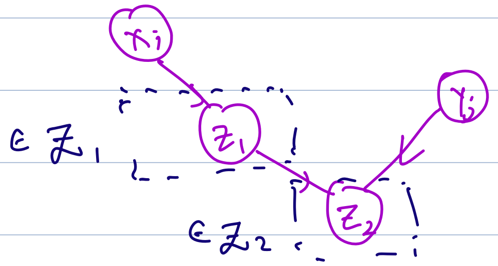
∴ By contradiction, parent(W) $\notin \mathcal{Y}$.

ⓑ For any $W \in \mathcal{Y} \cup \mathcal{Z}_2$, parent(W) $\subseteq \mathcal{Y} \cup \mathcal{Z}$

i.e., parent(W) $\notin \mathcal{X}$

Why? ↑ By definition of $\mathcal{Z}_2$

Aside:



Note that this is okay, because we are conditioning on BOTH (Z_1, Z_2). Thus, Z_1 blocks the path between X_1 and Y despite Z_2 being a collider.

Putting thing together:

$$P(\mathcal{X}, \mathcal{Y}, \mathcal{Z}) = \prod_{w \in \mathcal{X} \cup \mathcal{Y} \cup \mathcal{Z}} P(w \mid PA_w)$$

$$= \underbrace{\prod_{w \in \mathcal{X} \cup \mathcal{Z}_1} p(w|PA_w)}_{\phi_1(\mathcal{X}, \mathcal{Z})} \cdot \underbrace{\prod_{w \in \mathcal{Y} \cup \mathcal{Z}_2} P(w|PA_w)}_{\phi_2(\mathcal{Y}, \mathcal{Z})}$$

from ⓐ $(PA_w \notin \mathcal{Y})$ from ⓑ $(PA_w \notin \mathcal{X})$

$\Longrightarrow$ From Lemma 1, $\mathcal{X} \perp\!\!\!\perp \mathcal{Y} \mid \mathcal{Z}$ $\square$

Theorem 2 (Global Markov Property): Let X and Y be variables in a DAG, and $\mathcal{Z}$ any set of nodes s.t.

$$X \perp\!\!\!\perp_G Y \mid \mathcal{Z}. \quad \text{Then} \quad X \perp\!\!\!\perp Y \mid \mathcal{Z}$$

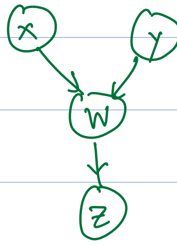
i.e., (d-separation) + (DAG) $\Longrightarrow$ Conditional independence.

Proof: $G = (V, E)$ is the DAG, $V = \{X_1, X_2, \ldots, X_d\}$.

Without loss of generality, assume that $V = \text{ancestor}\{\{X, Y\} \cup \mathcal{Z}\}$.

Why: We can prune out all other descendents

Example to keep in mind when reasoning through this proof:



recursively as in Lemma 3, and work with the resulting DAG.

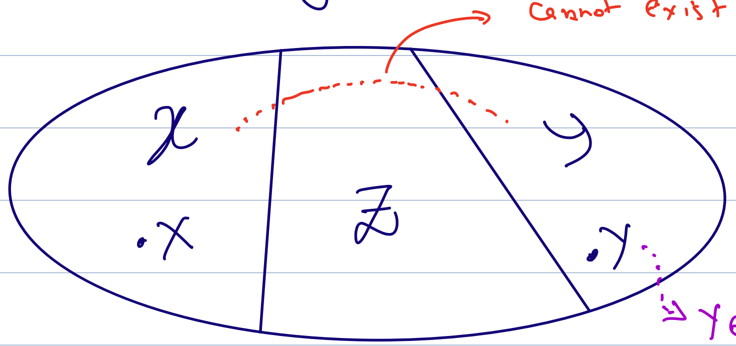
Let $\mathcal{X} = \{X_i : X_i \not\perp\!\!\!\perp_G X \mid \mathcal{Z}\}$

i.e., X_i are not d-seperated from X given $\mathcal{Z}$.

$\mathcal{Y} = V \setminus \mathcal{X} \cup \mathcal{Z}$

(i.e., all remaining nodes)

unblocked paths cannot exist



$Y \in \mathcal{Y}$ by defn of $\mathcal{X}$.

Claim: $\mathcal{X} \perp\!\!\!\perp_G \mathcal{Y} \mid \mathcal{Z}$, and $(\mathcal{X}, \mathcal{Y}, \mathcal{Z})$ forms a partition of V.

immediately holds by construction

For any $X_i \in \mathcal{X}$, and after conditioning on $\mathcal{Z}$, $\exists$ unblocked path between X_i and X (by defn of $\mathcal{X}$). $\therefore$ There <u>cannot</u> exist any unblocked path between X_i and $\mathcal{Y}$. This is because by definition of Y and $\mathcal{Y}$, all paths between X and $\mathcal{Y}$ are blocked.

$X \not\perp\!\!\!\perp_d Y \mid Z$

$\mathcal{X} = \{X, W\}$
$\mathcal{Z} = \{Z\}$
$\mathcal{Y} = \{Y\}$.

Observe above that X, W are not d-separated given Z.

If $\exists$ unblocked path from X_i and $\mathcal{Y}$, we can concatenate $X \overset{\text{unblocked path}}{\text{---}} X_i$ and $X_i \text{---} \mathcal{Y}$ to find an unblocked path from $X \text{---} \mathcal{Y}$, which is a contradiction.

$\therefore$ from Theorem 1, $\mathcal{X} \perp\!\!\!\perp \mathcal{Y} \mid \mathcal{Z}$. $\longrightarrow$ ①.

(The above argument is the crux of the proof).

We now need to show that ① $\Longrightarrow X \perp\!\!\!\perp Y \mid \mathcal{Z}$.

This follows by marginalizing out all nodes in $\mathcal{X}$ and $\mathcal{Y}$ other than X, Y. Details below:

$$\mathcal{X}' = \mathcal{X} \setminus \{X\} \qquad \mathcal{Y}' = \mathcal{Y} \setminus \{Y\}$$

$$A_i = \{X_i = x_i\}, \quad i = 1, 2, \cdots d$$

(Note: for notation, I am writing the proof for discrete; similar argument works with continuous)

$$x' = \{x_i : i \in \text{index-set of } \mathcal{X} \setminus \{X\}\}$$
$$y' = \{y_j : j \in \text{index-set of } \mathcal{Y} \setminus \{Y\}\}$$
$$z = \{z_k : k \in \text{index-set of } \mathcal{Z}\}$$

From Lemma 1 (characterization of CI),

$$x \perp\!\!\!\perp y \mid z \iff \exists\ \phi_1, \phi_2 \text{ s.t.}$$

$$P\left(\bigcap_{i=1}^{d} A_i\right) = \phi_1(\overbrace{x, x'}^{x}, z)\, \phi_2(\overbrace{y, y'}^{y}, z)$$

$$\therefore\ P\left(\{X = x\} \cap \{Y = y\} \cap \bigcap_{\{k : X_k \in Z\}} \{X_k = x_k\}\right)$$

$$= \sum_{x'} \sum_{y'} \phi_1(x, x', z)\, \phi_2(y, y', z)$$

$$= \sum_{x'} \phi_1(x, x', z) \sum_{y'} \phi_2(y, y', z)$$

$$= \tilde{\phi}_1(x, z)\, \tilde{\phi}_2(y, z)$$

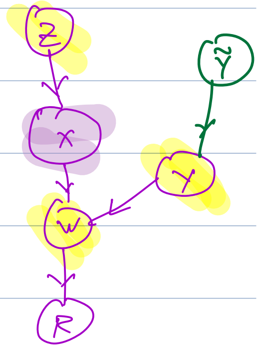
$$\implies X \perp\!\!\!\perp Y \mid Z$$

▨

Markov Blanket

Defn: The Markov Blanket for a node X consists of:

1. parent(X)
2. children(X)



3. parents-of-children(X)

Corollary 1: Let $\mathcal{B}$ be the Markov Blanket of X, and $\mathcal{Y} = V \setminus \{X\} \cup \mathcal{B}$. Then,

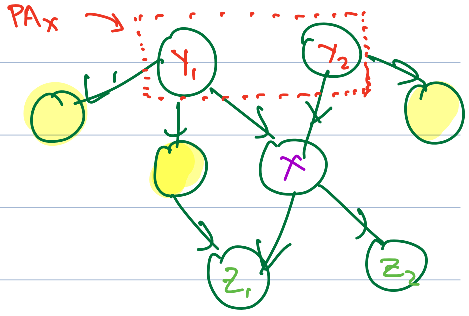
$$X \perp\!\!\!\perp \mathcal{Y} \mid \mathcal{B}.$$

For node X, we need to include Y in its Markov Blanket. This is because when conditioning on W (child of X), X and Y become unblocked

Pf: $\mathcal{B}$ d-separates X from any other node. The result immediately follows from Theorem 2 . $\square$

Corollary 2: (Local Markov Property)

Any node X in the DAG is conditionally indep. of all its non-descendents, given its parents.



X is cond. indep. of the ○ nodes, given Y_1, Y_2.

Proof: We need to show that

$$X \perp\!\!\!\perp_G \mathcal{R} \mid PA_X$$

PA_X ⟶ parents of X.

$\mathcal{R}$ ⟶ set of non-descendent nodes

i.e., show that any path between X and $\mathcal{R}$ is blocked. This follows because parents block paths going "upwards" and children (or their children somewhere downstream) become colliders and block paths to non-descendants "downwards".

Details: Consider any path between X and $R \in \mathcal{R}$. Let Z be a neighbor of X on this path. If Z is a parent(X), the conditioning on PA_X blocks this path.

If Z is not a parent (i.e., a child), then either Z is a collider, or some descendant of Z is a collider. ($\because$ This node is not in PA_X, and hence we ~~are~~ not conditioning on it).

∎

Theorem 3 (Markov Property)

DAG $\mathcal{G} = (V, E)$ and $p(\cdot)$ the joint dist associated with it:

(i) Global Markov Property: $\mathcal{X} \perp\!\!\!\perp_{\mathcal{G}} \mathcal{Y} \mid \mathcal{Z}$

$$\Longrightarrow \mathcal{X} \perp\!\!\!\perp \mathcal{Y} \mid \mathcal{Z}$$

(ii) Local Markov Property: $X \perp\!\!\!\perp \{\text{non-descendants}\} \mid PA_X$

(iii) Markov factorization Property:

$$p(x_1, \dots, x_d) = \prod_{j=1}^{d} p(x_i \mid PA_j)$$

Then, (i) $\iff$ (ii).

Further suppose that $\exists$ a product measure over the nodes $\mu = \bigotimes_{v \in V} \mu_v$ s.t. $p(\cdot)$ is absolutely continuous w.r.t. μ.

Then (i) $\iff$ (ii) $\iff$ (iii)

(Proof: see Independence Properties of Directed Markov Fields, Lauritzen et. al., Networks 1990).

Markov Equivalence of DAGs.

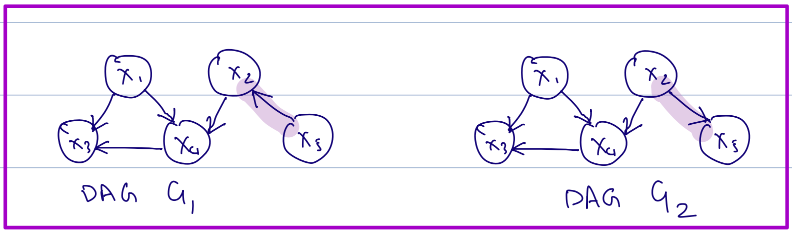
Given a directed graph $G = (V, E)$, let $\mathcal{M}(G)$ be the set of all distributions $p(\cdot)$ that have the

Markov property w.r.t $\mathcal{G}$, i.e.,

$$\mathcal{M}(\mathcal{G}) = \{ p : p(x_1, \cdots, x_d) \text{ has the Global Markov property w.r.t } \mathcal{G} \}$$

Definition: (Markov Equivalence of Graphs). DAGs $\mathcal{G}_1$ and $\mathcal{G}_2$ are Markov Equivalent if

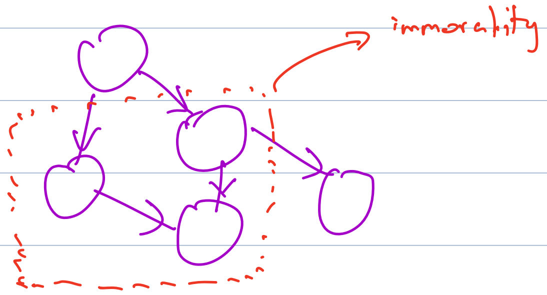
$$\mathcal{M}(\mathcal{G}_1) = \mathcal{M}(\mathcal{G}_2).$$



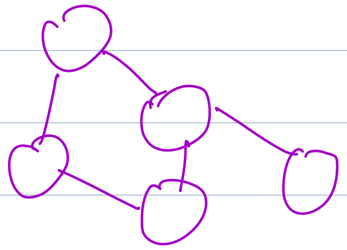
DAG $\mathcal{G}_1$ DAG $\mathcal{G}_2$

Please see Notes 2.b (Multiple factorizations) for additional discussion.

Defn: (Skeleton of DAG) The skeleton of a DAG $\mathcal{G}$ consists of the vertices along with the undirected edges.



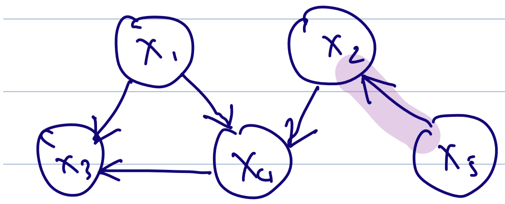
DAG $\mathcal{G}$



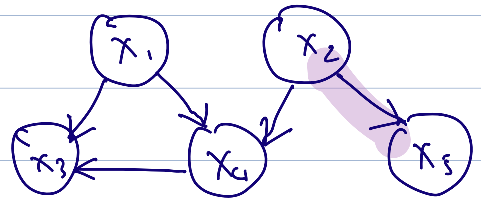
Skeleton of $\mathcal{G}$.

Defn: (Immorality) A collection of three nodes (X, Y, Z) form an immorality if $X \longrightarrow Y \longleftarrow Z$ (i.e., X and Z are parents of Y), but there is no edge between X and Z. (This is also called a unshielded collider)

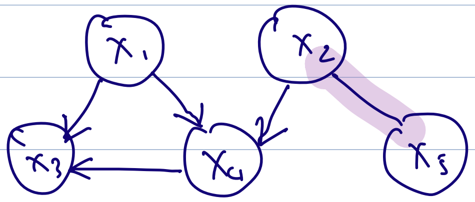
Example (Figure 6.4 in 'Elements of Causal Inference' book, pp. 103)



DAG G_1



DAG G_2



$$\text{CPDAG}(G_1) = \text{CPDAG}(G_2)$$

Graphs G_1 and G_2 above are Markov Equivalent.

Defn: (Markov Equivalence Class) The set of all DAGs that are Markov Equivalent to G is called its Markov Equivalence Class.

Lemma 4: G_1 and G_2 are Markov Equivalent

$$\Updownarrow$$

The graphs have the same skeleton and same imoralities.

aka V-structure aka unshielded collider

Completeted Partially Directed Acyclic Graph

Defn (CPDAG) Given a DAG $G = (V, E)$,

$\text{CPDAG}(G) = \{ (V, E')$: directed edge $e \in E'$ iff all members of the Markov Equivalence of G have the same directed edge; all other edges $e \in E$ are represented by undirected edges $\}$

(Causal) Minimality: A dist. $p(\cdot)$ is (causally) minimal w.r.t a DAG G if it is globally Markov w.r. G, but not any proper subset of G.

Remark: Causal minimality intuitively means that for each node, all its parents on G are active, in the sense that if we leave out that parent, CI relations for that node will not be true.

Prop 1 (6.36 in text) $(X_1, \cdots, X_d)$ associated with $G = (V, E)$. $p(\cdot)$ is abs. cont. w.r.t a product measure (see Theorem 3).

$$p \text{ causally minimal w.r.t. } G$$

$$\Updownarrow$$

$$\forall\ X_j \in V,\quad Y \in PA_j\ ,\quad X_j \not\perp\!\!\!\perp Y \mid PA_j \setminus \{Y\}.$$

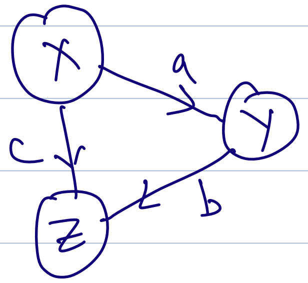
Faithfulness: $p(\cdot)$ is faithful to $G = (V, E)$ if

$$X_i \perp\!\!\!\perp X_j \mid Z \implies X_i \perp\!\!\!\perp_G X_j \mid Z$$

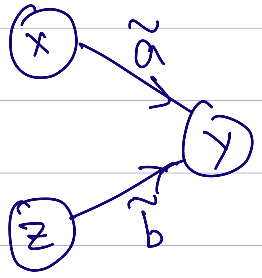
i.e., the converse of the Global Markov Property holds.

Remark: faithfulness is not always true!

Example 6.34 in text



G_1



G_2

$X = N_1$
$Y = aX + N_2$
$Z = cX + bY + N_3$.

$X = N_1$
$Y = \tilde{a}X + \tilde{b}Z + N_2$
$Z = N_3$

N_1, N_2, N_3 indep., $N_i \sim N(0,1)$.

Suppose $c + ab = 0$. Then, the two paths to Z in G_1 "cancel" each other out, meaning that $X \perp\!\!\!\perp Z$. However $X \not\!\perp\!\!\!\perp_G Z$.

Faithfulness violation causes us to be unable to distinguish (using even infinite number of samples drawn from $p(\cdot)$) between G_1 and G_2.

Prop 2: Let P_X Markovian w.r.t G. Then:

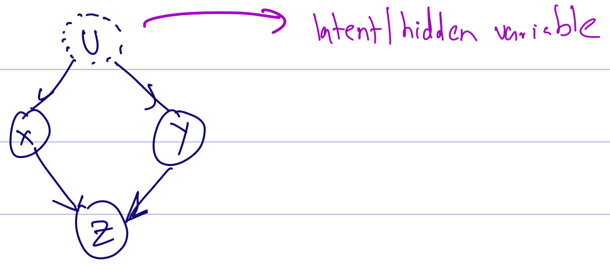
P_X faithful to G $\Longrightarrow$ P_X causally minimal w.r.t. G.

(see Prop 6.35 in text for proof).

Miscellaneous Remarks

Remark 1: Almost always in these notes (there are a few exceptions), we will assume that the entire DAG is visible, i.e., there are no latent / hidden variables.

If there are latent variables, they will be represented by a dotted circle:



This means that the joint distribution is given by

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

BUT we can only observe $p(x,y,z)$.

In general, we know that DAGs are not closed under marginalization, meaning that $p(x,y,z)$ need not have a Markov factorization, that encodes the independencies under the original DAG. (See Figure 1 in Silva and Ghahramani, ref below).

In this case, we need other graphical models that are closed under marginalization and/or conditioning. Some structures are MC-DAGs (Koster 2002), mDAGs (Evans 2015), etc. We are not going to study these structures. Please see refs below for discussion:

The hidden life of latent variables: Bayesian learning with mixed graph models, R. Silva and Z. Ghahramani, JMLR 2009.

Graphs for margins of Bayesian networks, R. Evans, 2015.
arXiv:1408.1809v2

Causality, J. Pearl, 2009.

Remark 2: Faithfulness is a strong assumption.
From the example above, it seemingly looks

like a mild assumption, as the example corresponds to a "zero measure" set of weights (i.e., set of $\{(a,b,c) \in \mathbb{R}^3 : c + ab = 0\}$ is a zero-Lebesgue measure set).

However, faithfulness is a strong assumption in practice. As shown in [a] below, the volume of SEMs that are "close" to ones with faithfulness violations is a large constant fraction of all linear SEMs. Thus, in a finite sample setting, distinguishing between CI and "noise" due to samples is difficult.

[a] C. Uhler, G. Raskutti, P. Buhlmann and B. Yu, Geometry of faithfulness assumption in causal inference, Annals of Statistics, 2013.