An Introduction to Graphical Models

Miles Bakenhus

2023-04-18

1 Notation and Basic Definitions

Notation:

- P: a probability distribution
- p: PDF or PMF associated with P. For discrete models $p \in \Delta_{r-1}$ for some $r \in \mathbb{Z}_{>0}$.
- $G = (V, E)$: a graph with vertex set V and edge set E (may be directed or undirected)
- $N(v)$: for G with $v \in V$, this is the neighborhood of v.
- $[n] = \{1, 2, \dots, n\}$
- $\mathcal{X}$ is the set of all possible values of random vector $X = (X_v \mid v \in V)$.

Definitions

Definition 1.1 (Graphs)**.** A *graph* $G = (V, E)$ is a set of vertices V and a set of edges E such that edges $(u, v) \in E$ are ordered pairs of vertices $u, v \in V$. Some graphs have a multi-set for E (called *multi-graphs*), however we will ignore these and focus on the following two types of graphs:

- An *undirected graph* has edges such that if $(u, v) \in E$ then $(v, u) \in E$. We say that two vertices u, v that share an edge in G are *adjacent* and write this $u \sim v$. Similarly if two vertices are not adjacent we write $u \nsim v$
- A *directed graph* distinguishes between edges (u, v) and (v, u). For adjacent vertices u, v, we denote $(u, v) \in E$ by $u \to v$ and $(v, u) \in E$ by $u \leftarrow v$. If $u \to v$ we say that u is a *parent* of v and that v is a *child* of u.

Definition 1.2 (Walks, Paths, and Cycles)**.**

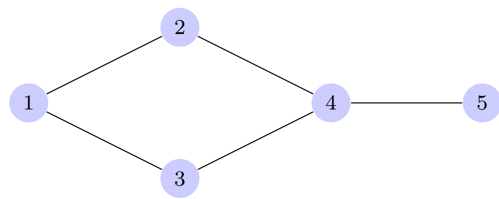
- For a graph $G = (V, E)$ and vertices $u, v \in V$, a *walk* of length k, is a sequence of vertices in G:

  $$u = v_0, v_1, \dots, v_k = v$$

  such such that $v_{i-1} \sim v_i$ for $i \in [k]$.
- A *path* is a walk such that all of the vertices in the sequence are distinct.
- A *cycle* is a walk $(v_0, \dots, v_k)$ such that $(v_1, \dots, v_k)$ is a path, and $v_0 = v_k$.
- A path (or cycle) $(v_0, v_1, \dots, v_k)$ on a directed graph is called *directed* whenever $v_{i-1} \to v_i$ for $i \in [k]$.

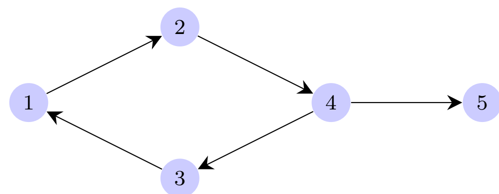
Definition 1.3 (Directed Acyclic Graphs)**.** A *directed acyclic graph (DAG)* is a directed graph that contains no directed cycles.

Example 1.1. For the following undirected graph:



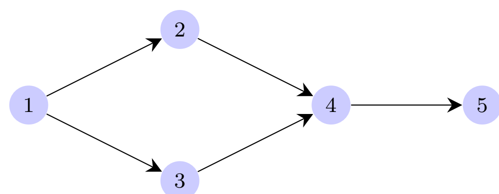
- $(1, 2, 4, 5, 4, 3)$ is a walk from vertex 1 to vertex 3.
- $(1, 2, 4, 3)$ is a path from vertex 1 to vertex 3.
- $(1, 2, 4, 3, 1)$ is a cycle

Example 1.2. For the following directed graph:



- $(1, 2, 4, 5)$ is a directed path from 1 to 5.
- $(1, 3, 4, 5)$ is a path from 1 to 5 but is not a directed path.
- $(1, 2, 4, 3, 1)$ is a directed cycle – note: this means G is **not** a DAG.
- $(1, 3, 4, 2, 1)$ is a cycle but not a directed cycle.

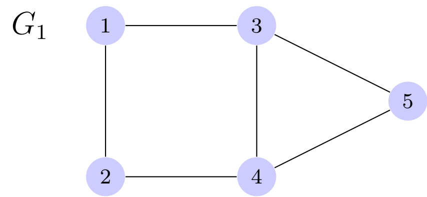
Example 1.3. The following is an example of a DAG:

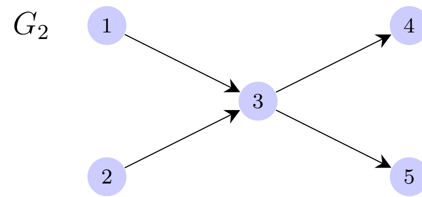


Notice how it contains cycles: $(1, 2, 4, 3, 1)$, $(1, 3, 4, 2, 1)$, but neither of these are directed.

2 Describing Graphical Models

Graphical models are a family of statistical models that describes the dependencies for a joint distribution P of the random variables $X = (X_v : v \in V)$, using the structure of a graph $G = (V, E)$. For example, if $X = (X_1, \dots, X_5)$ then the following graphs G_1 and G_2 could be potential graphical models for P:





Notice G_1 is undirected and G_2 is directed. These illustrate two major classes of graphical models that are commonly used. Edges and paths in a graphical model correspond to dependence relationships between random variables. This allows us to use the notion of *separation* for undirected graphs, and *d-separation* for directed graphs, to describe conditional independence relationships based on graph structure.

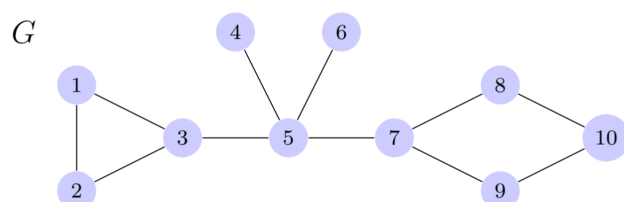
2.1 Undirected Models

Question: How do we determine the conditional independence statements described by an undirected model?

Suppose $G = (V, E)$ is an undirected graphical model for $X = (X_v : v \in V)$. Edges $(u, v) \in E$ correspond to dependence between X_u and X_v. Similarly, for two distinct vertices $u, v \in V$ where $u \nsim v$, dependence between X_u and X_v can be expressed using uv-paths in G. From the contraposition, if G models the independence relationship $X_u \perp\!\!\!\perp X_v$ then there should be no uv-path in G. We can extend this notion of paths and dependence to model conditional independence statements, using *separating sets* of vertices.

Definition 2.1 (Separating Set)**.** For a graph $G = (V, E)$, and distinct $u, v \in V$ where $u \nsim v$, the set $C \subseteq V \backslash \{u, v\}$, is a *separating set* for u and v if every uv-path contains at least one vertex in S. For disjoint sets of vertices $A, B \subseteq V$ where $a \nsim b$ for every $a \in A$ and every $b \in B$, the set $C \subseteq V \backslash (A \cup B)$ is a separating set for A and B if C is a separating set for every pair of vertices $a \in A$ and $b \in B$.

Example 2.1. Consider the graph $G = (V, E)$ below:



- For vertices $2, 8 \in V$, $2 \nsim 8$ and the following are some of the separating sets for these vertices:

$$\{5\}, \quad \{4, 5, 6\}, \quad \{1, 3, 7, 9, 10\}, \quad \{1, 3, 4, 5, 6, 7, 9, 10\}$$

- The disjoint sets $A = \{1, 2, 3\}$, $B = \{7, 8, 9, 10\}$, have only the following separating sets:

$$\{5\}, \quad \{4, 5\}, \quad \{5, 6\}, \quad \{4, 5, 6\}$$

- Sets of vertices may not necessarily give a connected induced subgraph. The disjoint sets $A = \{4, 6, 7\}$, $\{3, 10\}$ a minimal separating set is $C = \{5, 8, 9\}$.

Separating sets give us the means for describing conditional independence statements from a graphical model. Similar to how independence corresponds disconnected vertices in the graph, conditional independences correspond to sets of vertices and their separating sets.

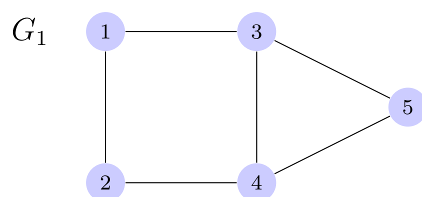
2.1.1 Markov Properties

The Markov properties for undirected graphical models give us a method for translating the structure of a graph into conditional independence statements. There are three Markov properties for undirected models. From definition 13.1.1 in (Sullivant 2018):

Definition 2.2 (Undirected Markov Properties)**.** For an undrected model $G = (V, E)$:

- The *pairwise* Markov property for undirected models associated to G gives the set of all conditional independence statements $X_v \perp\!\!\!\perp X_u \mid X_{V\setminus\{u,v\}}$ for $u \nsim v$ in G. This set is denoted $\text{pairwise}(G)$.
- The *local* Markov property for undirected models associated to G gives the set of all conditional independence statements $X_v \perp\!\!\!\perp X_{V\setminus(N(v)\cup\{v\})} \mid X_{N(v)}$ for all $v \in V$. This set is denoted $\text{local}(G)$.
- The *global* Markov property for undirected models associated to G gives the set of all conditional independence statements $X_A \perp\!\!\!\perp X_B \mid X_C$ for all disjoint $A, B, C \subseteq V$ where A and B share have no edges between them and C is a separating set for A and B. This set is denoted $\text{global}(G)$.

Example 2.2. Consider the graph G below:



- The pairwise Markov property gives one conditional independence statement for each pair of non-adjacent vertices. There are 4 such pairs in G: $\{1,5\}$, $\{1,4\}$, $\{2,3\}$, $\{2,5\}$. These give the following conditional independence statements:

$$\text{pairwise}(G) = \left\{ \left(X_1 \perp\!\!\!\perp X_5 \mid X_{\{2,3,4\}}\right), \left(X_1 \perp\!\!\!\perp X_4 \mid X_{\{2,3,5\}}\right), \left(X_2 \perp\!\!\!\perp X_3 \mid X_{\{1,4,5\}}\right), \left(X_2 \perp\!\!\!\perp X_5 \mid X_{\{1,3,4\}}\right) \right\}$$

- The local Markov property gives a conditional independence statement for each vertex in the graph, conditioning on the neigborhoods:

$$\text{local}(G) = \left\{ \left(X_1 \perp\!\!\!\perp X_{\{4,5\}} \mid X_{\{2,3\}}\right), \left(X_2 \perp\!\!\!\perp X_{\{3,5\}} \mid X_{\{1,4\}}\right), \left(X_3 \perp\!\!\!\perp X_2 \mid X_{\{1,4,5\}}\right) \left(X_4 \perp\!\!\!\perp X_1 \mid X_{\{2,3,5\}}\right)), \left(X_5 \perp\!\!\!\perp X_{\{1,2\}} \mid X_{\{3,4\}}\right) \right\}$$

- The set $\text{global}(G)$ is generally more difficult to compute because it considers disjoint sets of vertices with no edges between them and their separating sets. For the global properties computed in Macaulay2 see the code example.

<u>**Question:**</u> We can see in the example that there is some overlap in the statements generated by these properties. Do some properties imply others?

It is relatively straightforward to see that $\text{global}(G) \Rightarrow \text{local}(G), \text{pairwise}$. For any vertex $v \in V$ we can take $A = \{v\}$ and $B = V\setminus(N(v) \cup \{v\})$. Clearly, these are disjoint sets with no adjacent vertices between them, and they are separated by the set $C = N(v)$. Then if we have all of $\text{global}(G)$ we can recover the statements in $\text{local}(G)$. Similarly, for non-adjacent $u, v \in V$, we can let $A = \{v\}$, $B = \{u\}$ and $C = V\setminus\{u, v\}$ to recover $\text{pairwise}(G)$. Does $\text{local}(G) \Rightarrow \text{pairwise}(G)$?

Macaulay2 code example:

```
# The following creates a temporary Macaulay2 Script and runs it.
# Everything between the two EOFs is Macaulay2 code.
cat > tmp.m2 << EOF
-- Loads the Graphical Models Package
loadPackage "GraphicalModels";

-- Define the graph in our example
G = graph { {1,{2,3}}, {2,{4}}, {3,{4,5}}, {4,{5}} }

print"Pairwise Conditional Independence Statements:"
-- Computes pairwise statements
pairCIs = pairMarkov G
-- Prints each statement
for i from 0 to length(pairCIs)-1 do (print pairCIs_i)

print"\nLocal Conditional Independence Statements:"
-- Computes local statements
localCIs = localMarkov G
for i from 0 to length(localCIs)-1 do (print localCIs_i)
print"These are different from our local statements (Hint: prop 4.1.4 weak union)"

print"\nGlobal Conditional Independence Statements:"
-- Computes global statements
globalCIs = globalMarkov G
for i from 0 to length(globalCIs)-1 do (print globalCIs_i)
print"These are the same as the local statements but shouldn't this include all"
print"disjoint sets with no adjacent vertices? (Hint: weak union again)"
EOF
M2 --script tmp.m2
rm tmp.m2
```

```
## Pairwise Conditional Independence Statements:
## {{1}, {4}, {5, 2, 3}}
## {{1}, {5}, {4, 2, 3}}
## {{2}, {3}, {4, 5, 1}}
## {{2}, {5}, {4, 1, 3}}
##
## Local Conditional Independence Statements:
## {{1}, {4, 5}, {2, 3}}
## {{1, 2}, {5}, {4, 3}}
## {{2}, {3, 5}, {4, 1}}
## These are different from our local statements (Hint: prop 4.1.4 weak union)
##
## Global Conditional Independence Statements:
## {{1}, {4, 5}, {2, 3}}
## {{1, 2}, {5}, {4, 3}}
## {{2}, {3, 5}, {4, 1}}
## These are the same as the local statements but shouldn't this include all
## disjoint sets with no adjacent vertices? (Hint: weak union again)
```

To see how local(G) $\Rightarrow$ pairwise(G) we need proposition 4.1.4 from (Sullivant 2018):

Proposition 2.1 (Conditional Independence Axioms)**.** Let $A, B, C, D \subseteq V$ be pairwise disjoint subsets. Then

(i) (*symmetry*) $X_A \perp\!\!\!\perp X_B \mid X_C \implies X_A \perp\!\!\!\perp X_B \mid X_C$;

(ii) (*decomposition*) $X_A \perp\!\!\!\perp X_{B\cup D} \mid X_C \implies X_A \perp\!\!\!\perp X_B \mid X_C$;

(iii) (*weak union*) $X_A \perp\!\!\!\perp X_{B\cup D} \mid X_C \implies X_A \perp\!\!\!\perp X_B \mid X_{C\cup D}$;

(iv) (*contraction*) $X_A \perp\!\!\!\perp X_B \mid X_{C\cup D}$ and $X_A \perp\!\!\!\perp X_D \mid X_C \implies X_A \perp\!\!\!\perp X_{B\cup D} \mid X_C$.

Using the weak union property in particular, we prove the claim local(G) $\Rightarrow$ pairwise(G):

Proof. Suppose the conditional independence statements in local(G) are satisfied for some joint distribution P on X. Then for each $v \in V$, $X_v \perp\!\!\!\perp X_{V\setminus(N(v)\cup\{v\})} \mid X_{N(v)}$. Notice that whenever $u \nsim v$ for $u \in V$ then

$$V\setminus(N(v) \cup \{v\}) = \{u\} \cup [V\setminus(N(v) \cup \{u, v\})]$$

Let $A = \{v\}$, $B = \{u\}$, $C = N(v)$ and $D = V\setminus(N(v) \cup \{u, v\})$. Then from the weak union property

$$X_A \perp\!\!\!\perp X_{B\cup D} \mid X_C \;\Rightarrow\; X_A \perp\!\!\!\perp X_B \mid X_{C\cup D} \Rightarrow X_v \perp\!\!\!\perp X_u \mid X_{N(v)\cup[V\setminus(N(v)\cup\{u,v\})]}$$

Since

$$N(v) \cup [V\setminus(N(v) \cup \{u, v\})] = [N(v) \cup (V\setminus N(v))] \setminus\{u, v\} = V\setminus\{u, v\}$$

this gives the pairwise conditional independence statement:

$$X_v \perp\!\!\!\perp X_u \mid X_{N(v)\cup[V\setminus(N(v)\cup\{u,v\})]} \;\Rightarrow\; X_v \perp\!\!\!\perp X_u \mid X_{V\setminus\{u,v\}}$$

□

We now have the following fact for distributions modelled by undirected graphical models:

$$\text{global}(G) \Rightarrow \text{local}(G) \Rightarrow \text{pairwise}(G)$$

Question: Are these properties ever equivalent?

Recall proposition 4.1.5, from (Sullivant 2018):

Proposition 2.2 (Intersection Axiom)**.** If the join pdf $f(x) > 0$ (or pmf $p(x) > 0$ in the discrete case) for all $x \in \mathcal{X}$, then for disjoint subsets $A, B, C, D \subseteq V$:

$$X_A \perp\!\!\!\perp X_B \mid X_{C\cup D} \text{ and } X_A \perp\!\!\!\perp X_C \mid X_{B\cup D} \implies X_A \perp\!\!\!\perp X_{B\cup C} \mid X_D$$

Using the intersection axiom it is possible to show Theorem 13.1.4 from (Sullivant 2018):

Theorem 2.1. *If the random vector X has joint distribution* P *that satisfies the intersection axiom then* P *obeys the pairwise Markov property fo the undirected graph G if and only if it obeys the global Markov property for the graph G. In particular, if $p(x) > 0$ for all x, then the pairwise, local, and global Markov properties are equivalent.*

Proof Sketch. We've seen how global(G) $\Rightarrow$ pairwise(G) so we only need to show pairwise(G) $\Rightarrow$ global(G) when the intersection axiom applies to P.

Suppose that intersection axiom holds for the join distribution P of the random variable X, and satisfies the conditional independence statements in pairwise(G) for some $G = (V, E)$. Let $A, B, C \subseteq V$ be disjoint subsets such that C separates A and B in G. We can assume $A \cup B \cup C = V$, since we can always find disjoint supersets $A', B', C' \subseteq V$ such that $A \subseteq A'$, $B \subseteq B'$, and $C \subseteq C'$, and $A' \cup B' \cup C' = V$. Then applying the decomposition axiom from proposition 4.1.4 allows us to recover the original sets in the global conditional independence statements.

For each pair $a \in A$ and $b \in B$, $a \nsim b$ since C separates A and B, so the pairwise statements hold: $X_a \perp\!\!\!\perp X_b \mid X_{V \backslash \{a,b\}}$. Without loss of generality, assume B has more than one element (otherwise we would have started with a pairwise statement), such that there exist distinct $b, b' \in B$. Clearly $a \nsim b'$ such that we have two parwise statements:

$$X_a \perp\!\!\!\perp X_b \mid X_{V \backslash \{a,b\}} \text{ and } X_a \perp\!\!\!\perp X_{b'} \mid X_{V \backslash \{a,b'\}}$$

Denote the following sets:

$$A_0 = \{a\}, \quad B_0 = \{b\}, \quad C_0 = \{b'\}, \quad D_0 = V \backslash (\{a\} \cup \{b, b'\})$$

Then

$$V \backslash \{a, b\} = \{b'\} \cup [V \backslash (\{a\} \cup \{b, b'\})] = C_0 \cup D_0$$
$$V \backslash \{a, b'\} = \{b\} \cup [V \backslash (\{a\} \cup \{b, b'\})] = B_0 \cup D_0$$

such that the follow statements hold:

$$X_{A_0} \perp\!\!\!\perp X_{B_0} \mid X_{C_0 \cup D_0} \text{ and } X_{A_0} \perp\!\!\!\perp X_{B_0} \mid X_{C_0 \cup D_0}$$

Applying the intersection gives:

$$X_{A_0} \perp\!\!\!\perp X_{B_0 \cup C_0} \mid X_{D_0} \Rightarrow X_a \perp\!\!\!\perp X_{\{b,b'\}} \mid X_{V \backslash (\{a\} \cup \{b,b'\})}$$

Then repeating this same procedure for every $b \in B$ gives the independence statement: $X_a \perp\!\!\!\perp X_B \mid X_{V \backslash (\{a\} \cup B)}$. We can also do this for all $a \in A$ to get $X_A \perp\!\!\!\perp X_B \mid X_{V \backslash (A \cup B)}$. Now since $A \cup B \cup C = V$ then $V \backslash (A \cup B) = C$, such that the global conditional independence statement $X_A \perp\!\!\!\perp X_B \mid X_C$ holds. $\square$

2.1.2 Summarizing Undirected Models

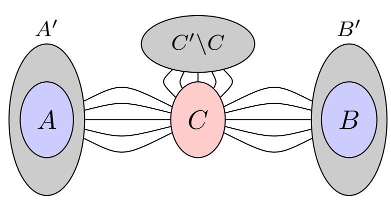
From Theorem 13.1.4 we have the key takeaway that $p(x) > 0$ for all $x \in \mathcal{X}$ implies that

$$\text{global}(G) \iff \text{local}(G) \iff \text{pairwise}(G)$$

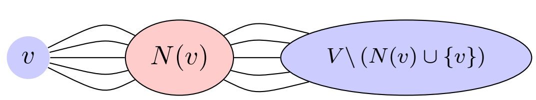
This will be important when we define parameterizations based on graphical models.

The following graphics summarize the Markov properties for undirected models $G = (V, E)$:

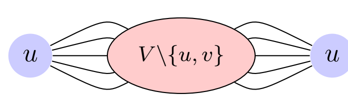
- **Global Markov Property**: For disjoint sets $A, B, C \subseteq V$ where A and B share no edges and C separates A and B, with supersets $A', B', C' \subseteq V$ such that $A \subseteq A'$, $B \subseteq B'$, and $C \subseteq C'$, and $A' \cup B' \cup C' = V$:



- **Local Markov Property**: For every vertex $v \in V$ with neighborhood $N(v)$:



- **Pairwise Markov Property**: For every pair of distinct vertices $u, v \in V$ where $u \nsim v$:



2.2 Directed Models

2.2.1 Motivation

<u>**Question:**</u> Why do we need directed models when we have undirected models?

Example 2.3. Consider a random contingency table consisting of counts of joint observations of two binary random variables X_1 and X_2:

$$U = \begin{pmatrix} U_{11} & U_{12} \\ U_{21} & U_{22} \end{pmatrix}$$

Let $T(U) = (U_{1+}, U_{2+}, U_{+1}, U_{+2})$ be a statistic for U that computes the marginals of the table. Suppose we sample values of U such that $U_{11} \perp\!\!\!\perp U_{12}$. Clearly U_{1+} depends on the values of U_{11} and U_{12} but is it true that $U_{11} \perp\!\!\!\perp U_{12} \mid U_{1+}$? In other words are U_{11} and U_{12} still independent if we know the rowsum? Using the probability chain rule:

$$\mathrm{P}(U_{11} = u_{11}, U_{12} = u_{12} \mid U_{1+} = u_{1+}) = \mathrm{P}(U_{11} = u_{11})\,\mathrm{P}(U_{12} = u_{12} \mid U_{11} = u_{11}, U_{1+} = u_{1+})$$

Since $U_{1+} = U_{11} + U_{12}$, then given $U_{11} = u_{11}$ and $U_{1+} = u_{1+}$ it must hold that

$$u_{1+} = u_{11} + U_{12} \Rightarrow U_{12} = u_{1+} - u_{11}$$

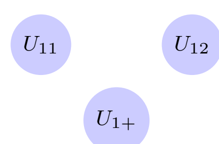
Then

$$\mathrm{P}(U_{12} = u_{12} \mid U_{11} = u_{11}, U_{1+} = u_{1+}) = \begin{cases} 0, & \text{when } u_{12} \neq u_{1+} - u_{11} \\ 1, & \text{when } u_{12} = u_{1+} - u_{11} \end{cases}$$

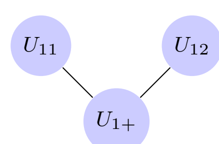
However, in general: $\mathrm{P}(U_{12} = u_{12} \mid U_{1+} = u_{1+}) \neq 0, 1$. Since

$$\mathrm{P}(U_{12} = u_{12} \mid U_{11} = u_{11}, U_{1+} = u_{1+}) \neq \mathrm{P}(U_{12} = u_{12} \mid U_{1+} = u_{1+})$$

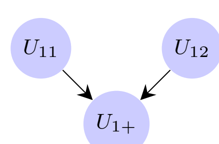
it means that $U_{11} \not\perp\!\!\!\perp U_{12} \mid U_{1+}$. There is not an undirected graphical model that models all fo the independence relationships: $U_{11} \perp\!\!\!\perp U_{12}$, $\{U_{11}, U_{12}\} \not\perp\!\!\!\perp U_{1+}$, and $U_{11} \not\perp\!\!\!\perp U_{12} \mid U_{1+}$, simultaneously. In the following model, $U_{11} \perp\!\!\!\perp U_{12}$ but $U_{11} \perp\!\!\!\perp U_{12} \mid U_{1+}$ and $\{U_{11}, U_{12}\} \perp\!\!\!\perp U_{1+}$



For this model, $\{U_{11}, U_{12}\} \not\perp\!\!\!\perp U_{1+}$ but $U_{11} \not\perp\!\!\!\perp U_{12}$ and $U_{11} \perp\!\!\!\perp U_{12} \mid U_{1+}$:



Indstead, a directed model can be used:



In the above, $U_{11} \perp\!\!\!\perp U_{12}$, $\{U_{11}, U_{12}\} \not\perp\!\!\!\perp U_{1+}$, and $U_{11} \not\perp\!\!\!\perp U_{12} \mid U_{1+}$, is are modeled using what is known as a *collider* (or *v-structure*).

2.2.2 d-Separation

While undirected graphs model dependence through path-connectedness between vertices and model conditional independence through separating sets, directed models use the slightly different concepts of *d-connectedness* and *d-separation*.

First we need to define some notation. For a DAG $G = (V, E)$ and vertex $v \in V$:

- $\mathrm{pa}(v) = \{u \in V : u \to v\}$ is the set of *parents* of v.
- $\mathrm{de}(v) = \{u \in V :$ there exists a directed vu-path$\}$ is the set of *descendants* of v.
- $\mathrm{nd}(v) = V \backslash (\{v\} \cup \mathrm{de}(v))$ is the set of *non-descendants* of v.
- $\mathrm{an}(v) = \{u \in V :$ there exists a directed uv-path$\}$ is the set of *ancestors* of v.

Example 2.4. In directed graphs below, we have the following relationships:

1. $u \in \mathrm{pa}(v)$, $u \in \mathrm{nd}(v)$, and $u \in \mathrm{an}(v)$
2. $u \in \mathrm{an}(v)$, $u \in \mathrm{nd}(v)$, and $u \notin \mathrm{pa}(v)$
3. $u \in \mathrm{nd}(v)$, $u \notin \mathrm{pa}(v)$, and $u \notin \mathrm{an}(v)$ (Note: this is a collider!)
4. $u \in \mathrm{de}(v)$

(1) (2) (3) (4)

Using these sets we can define notions of *d-connectedness* and *d-separation*. From definition 13.1.7 in (Sullivant 2018):

Definition 2.3 (d-connected)**.** For $u, v \in V$ and a set $C \subseteq V \backslash \{u, v\}$, u and v are *d-connected* by a given set C if there is an undirected path $\pi = (u, v_1, \dots, v_{k-1}, v)$ from u to v such that:

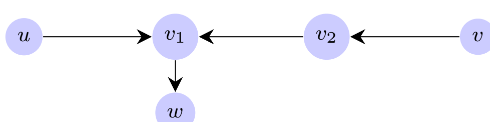
(i) all colliders on π are in $C \cup \mathrm{an}(C)$

(ii) no non-collider on π is in C.

This is somewhat abstract so consider the following example:

Example 2.5. For the DAG given bellow for which C's are u and v d-connected?

1. $C = \emptyset$
2. $C = \{v_1\}$
3. $C = \{w\}$
4. $C = \{w, v_2\}$

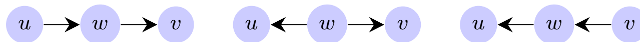


There is a single undirected uv-path to consider $\pi = (u, v_1, v_2, v)$:

1. u, v are d-separated (they are *not d-connected*) when $C = \emptyset$ because v_1 on π is a collider and is not in $C \cup \mathrm{an}(C)$.
2. u, v are d-connected when $C = \{v_1\}$, since v_1, the only collider of π, is included in $C \cup \mathrm{an}(C)$, and none of the non-colliders of π are included in C.
3. u, v are d-connected when $C = \{w\}$, since the set $\mathrm{an}(w) = \{u, v_1, v_2, v_3\}$ such that the collider $v_1 \in C \cup \mathrm{an}(C)$, and there are no non-colliders in C.
4. u, v are d-separated when $C = \{w, v_2\}$ because the non-collider $v_2 \in C$.

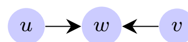
In practice it is often simpler to think of d-connectedness in terms of "active" and "inactive" paths in the DAG (Koller and Friedman 2009). In particular, we have the following scenarios for a DAG with verices u, v, w:

1. For the three directed paths



   for a given set C, these paths are "active" (d-connected) when $w \notin C$ and "inactive" when $w \in C$.

2. In the path (collider!):



   for a given set C, these paths are "inactive" (d-separated) when $\{w\} \cup \text{de}(w)$ is disjoint from C and "active" when $(\{w\} \cup \text{de}(w)) \cap C$ is non-empty. Notice that when C contains any descendants of w then $w \in \text{an}(C)$, so whenever a collider is on a path we are checking for d-separation, it is important to check $\text{de}(w) \cap C$ even when w is not included in C.

When deciding whether a uv-path is d-connected give C, we can look for any parts of the path that are inactive for C. If none are found then the whole path is active, and u and v are d-connected. Since colliders are active whenever their descendants are in C we should pay special attention to these vertices in the DAG. The notion of d-separation can be extended to disjoint sets of vertices A, B, C where there are no directed edges between A and B: C d-separates A and B whenever C d-separates all $a \in A$ from all $b \in B$.

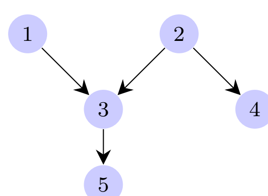
2.2.3 Directed Markov Properties

Using the concept of d-connectedness, we can define *directed Markov Properties* for DAGs $G = (V, E)$. From definition 13.1.9 in (Sullivant 2018) and (Lauritzen 2011):

Definition 2.4 (Directed Markov Properties)**.** For an undrected model $G = (V, E)$:

- The *directed pairwise* Markov property for undirected models associated to G gives the set of all conditional independence statements $X_u \perp\!\!\!\perp X_v \mid X_{(\text{an}(u)\cup\text{an}(v))\setminus\{u,v\}}$ where $u \nsim v$ in G. This set is denoted pairwise(G). Alternatively this can be defined as $X_u \perp\!\!\!\perp X_v \mid X_{(\text{nd}(u)\cup\text{nd}(v))\setminus\{u,v\}}$ for $u \nsim v$.
- The *directed local* Markov property for undirected models associated to G gives the set of all conditional independence statements $X_v \perp\!\!\!\perp X_{\text{nd}(v)\setminus\text{pa}(v)} \mid X_{\text{pa}(v)}$ for all $v \in V$. This set is denoted local(G).
- The *directed global* Markov property for undirected models associated to G gives the set of all conditional independence statements $X_A \perp\!\!\!\perp X_B \mid X_C$ for disjoint $A, B, C \subset V$ where A and B share no directed edges, such that C d-separates A and B. This set is denoted global(G).

Example 2.6. Given the DAG below, consider the pairwise, local conditional independence statements.



The pairwise statements from the DAG are (alternative definition statements in parentheses)

$$\begin{aligned}\text{pairwise}(G) = \quad &\big\{X_1 \perp\!\!\!\perp X_2 \text{ (or } X_1 \perp\!\!\!\perp X_2 \mid X_4), && X_1 \perp\!\!\!\perp X_4 \mid X_2 \text{ (or } X_1 \perp\!\!\!\perp X_4 \mid X_{2,3,5}),\\ &X_1 \perp\!\!\!\perp X_5 \mid X_{2,3} \text{ (or } X_1 \perp\!\!\!\perp X_5 \mid X_{2,3,4}), && X_2 \perp\!\!\!\perp X_5 \mid X_{1,3} \text{ (or } X_2 \perp\!\!\!\perp X_5 \mid X_{1,3,4}),\\ &X_3 \perp\!\!\!\perp X_4 \mid X_{1,2} \text{ (or } X_3 \perp\!\!\!\perp X_4 \mid X_{1,2,5}), && X_4 \perp\!\!\!\perp X_5 \mid X_{1,2,3}\big\}\end{aligned}$$

The local statements are:

$$\begin{aligned}\text{local}(G) = \quad &\big\{X_1 \perp\!\!\!\perp X_{2,4}, && X_3 \perp\!\!\!\perp X_4 \mid X_{1,2},\\ &X_4 \perp\!\!\!\perp X_{1,3,5} \mid X_2, && X_5 \perp\!\!\!\perp X_{1,2,4} \mid X_3\big\}\end{aligned}$$

Notice that $X_2 \perp\!\!\!\perp X_{\mathrm{nd}(2)\setminus \mathrm{pa}(2)} \mid X_{\mathrm{pa}(2)}$ was not included in the local statements. That is because this statement is $X_2 \perp\!\!\!\perp X_1$, which is implied by the first statement.

Unlike undirected models, theorem 13.1.11 in (Sullivant 2018) gives the equivalence $\mathrm{global}(G) \Leftrightarrow \mathrm{local}(G)$ without any assumptions about the distribution modeled by G. That said, directed models still need $p(x) > 0$ for all x such that the intersection axiom holds, in order to have equivalence between all 3 properties:

$$\mathrm{global}(G) \Leftrightarrow \mathrm{local}(G) \Leftrightarrow \mathrm{pairwise}(G)$$

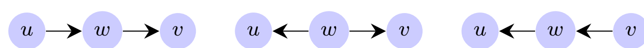
2.3 Uniqueness of Conditional Independence Statements

Given an undirected or a directed model we know how to find conditional independence statements using the Markov properties, for the modeled joint distribution $p(X)$ for $X = (X_v \mid v \in V)$.

Question: Are the sets of independence statements unique to a given graphical model?

For undirected models, $\mathrm{global}(G)$ includes $\mathrm{local}(G)$ which is defined by the vertex neighborhoods in a graph. Then if $\mathrm{global}(G) = \mathrm{global}(G')$ where $G = (V, E)$ and $G' = (V, E')$ then for $v \in V$, the neighborhood of v in G, is the same as the neighborhood of v in G'. Then both graphs have the same edges, meaning they are equivalent. In other words, for an undirected graph G, the set of conditional independence statements $\mathrm{global}(G)$ is unique. This is *not* true for directed models.

Example 2.7. Recall the three active paths from the earlier:



$G_1 = (V, E_1)$ $G_2 = (V, E_2)$ $G_3 = (V, E_3)$

Each of these DAGs has a different edge set: $E_1 = \{uw, wv\}$, $E_2 = \{wu, wv\}$, $E_3 = \{wu, vw\}$. However they all correspond to the same conditional independence statement: $X_u \perp\!\!\!\perp X_v \mid X_w$.

When two DAGs correspond to the same set of independence statements they are called **Markov Equivalent**. Theorem 13.1.13 from (Sullivant 2018) gives the conditions under which two DAGs are Markov equivalent:

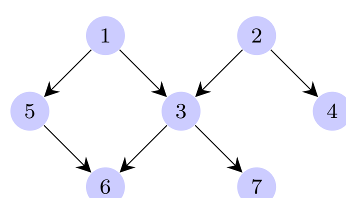
Theorem 2.2 (Markov Equivalence)**.** *Two DAGs $G_1 = (V, E_1)$ and $G_2 = (V, E_2)$ are Markov equivalent if and only if these two conditions are satisfied:*

(i.) *G_1 and G_2 have the same underlying undirected graph.*

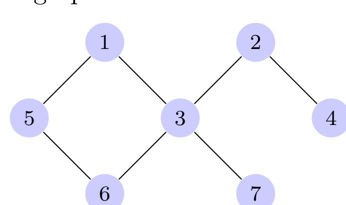
(ii.) *G_1 and G_2 have the same unshielded colliders: $u, v, w \in V$ such the induced subgraph of these three vertices is $u \to w \leftarrow v$, where $u \nsim v$.*

Example 2.8. Find DAGs that are Markov equivalent to $G = (V, E)$ below:

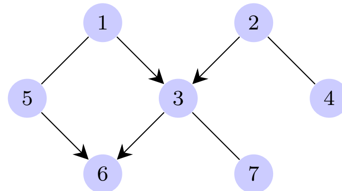
G



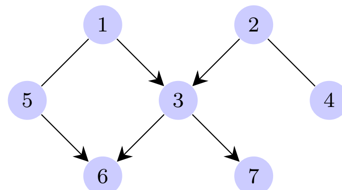
First consider the underlying undirected graph:



Any Markov equivalent graphs should have the same unshielded colliders, so adding those back into the graph gives:

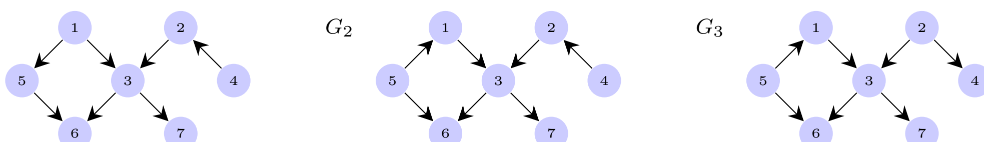


Since a directed edge $7 \to 3$ would create new colliders $1 \to 3 \leftarrow 7$ and $2 \to 3 \leftarrow 7$, there is only one possible orientation of this edge to maintain Markov equivalence:



The above graph is often called the *annotated skeleton* of G. This is a graph where the underlying undirected graph has directed edges showing the unshielded colliders of the original graph, and directed edges that are required to have a specific orientation to prevent new colliders from being created. The Markov equivalent graphs for G are DAGs where the remaining undirected edges of the annotated skeleton are directed in such a way that no new colliders are created. This gives the following graphs a Markov equivalent to G:

G_1 G_2 G_3



3 Parameterizations

3.1 Implicit Model Description

In the last section, we discussed how to define graphical models in terms of the Markov properties for both undirected and directed graphs. These Markov properties give a method for determining the conditional independence statements for a joint probability distribution of $X = (X_v \mid v \in V)$ modeled by a graph with vertex set V. Conditional independence statements implicitly characterize a joint probability distribution through factorization. For instance

$$\begin{array}{rcrcl} X \perp\!\!\!\perp Y & \Rightarrow & p(x,y) & = & p_X(x)p_Y(y) \\ Y \perp\!\!\!\perp Z \mid X & \Rightarrow & p(x,y,z) & = & p_X(x)p_{Y|X}(y \mid x)p_{Z|X}(z \mid x) \end{array}$$

The factorization for a modeled joint distribution gives a parametric description of the model in terms of these factors. In the independence model, we can write:

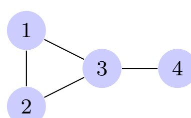
$$\mathrm{P}(X = i, Y = j) = p_X(i)p_Y(j) = \alpha_i\beta_j$$

for some parameters α_i and β_j that map to the factors $p_X(i)$ and $p_Y(j)$. In this section, the methods for describing the parametrizations for undirected and directed models will be discussed in further detail.

3.2 Undirected Parameterizations

Parameterizations for undirected models are described in terms of the *cliques* of the graph $G = (V, E)$. A clique is a set of vertices $C \subseteq V$ such that the induced sub graph G_C is complete, i.e. every vertex in C is adjacent to every other vertex in C. A clique C is called *maximal* in G whenever C is not contained in any other clique of G.

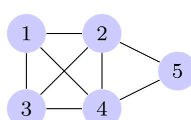
Example 3.1. For G_1:



The cliques of G are:

- The verices: $\{1\}, \{2\}, \{3\}, \{4\}$
- The edges: $\{1,2\}, \{1,3\}, \{2,3\}, \{3,4\}$
- The triangle: $\{1,2,3\}$

For G_2:



The set of vertices $\{1,2,3\}$ is a clique, but it is not maximal since it is contained in the clique $\{1,2,3,4\}$. The maximal cliques of G_2 are $\{1,2,3,4\}$ and $\{2,4,5\}$.

Denote the set of all maximal cliques of an undirected graph $G = (V, E)$ with $\mathcal{C}(G)$ and the state space of $X = (X_v \mid v \in V)$ by $\mathcal{X}$. For each clique $C \in \mathcal{C}$ denote the state space of $X_C = (X_v \mid v \in C)$ by $\mathcal{X}_C$. A continuous probability distribution P is said to factorize according to G if its pdf f can be written:

$$f(x) = \frac{1}{Z} \prod_{C \in \mathcal{C}(G)} \phi_C(x_C); \quad Z = \int_{\mathcal{X}} \prod_{C \in \mathcal{C}(G)} \phi_C(x_C) d\,\mathrm{P}$$

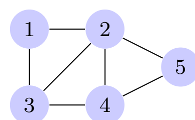
where ϕ_C (called a *potential function*) is a continuous function such that $\phi_C(x_C) \geq 0$ for all $x_C \in \mathcal{X}_C$.

Question: What does each part of this factorization mean?

- $\phi_C(x_C)$ are factors corresponding to each clique $C \in \mathcal{C}(G)$.
- $\prod_{C\in\mathcal{C}(G)} \phi_C(x_C)$ considers all of the factors according to the structure of G, defined by the maximal cliques.
- $\frac{1}{Z}$ is normalizing constant called the *partition function*. This is often difficult to compute, but is conceptually, just a way to make sure the factorization gives a probability distribution:

$$\int_{\mathcal{X}} f(x) d\,\mathrm{P} = \int_{\mathcal{X}} \frac{1}{Z} \prod_{C\in\mathcal{C}(G)} \phi_C(x_C) d\,\mathrm{P} = \frac{1}{Z} \int_{\mathcal{X}} \prod_{C\in\mathcal{C}(G)} \phi_C(x_C) d\,\mathrm{P} = \frac{1}{Z}(Z) = 1$$

Example 3.2. The undirected model G:



has the maximal cliques: $\mathcal{C}(G) = \{\{1,2,3\},\ \{2,3,4\},\ \{2,4,5\}\}$. The factorization has a potetential function for each of these cliques: $\phi_{1,2,3},\ \phi_{2,3,4}\ \phi_{2,4,5}$, such that

$$f(x_1, x_2, x_3, x_4, x_5) = \frac{1}{Z}\phi_{1,2,3}(x_1, x_2, x_3)\phi_{2,3,4}(x_2, x_3, x_4)\phi_{2,4,5}(x_2, x_4, x_5)$$

Question: When does a probability distribution factor according to a graph G?

The Hammersley-Clifford theorem gives the conditions under which a distribution factors. A proof of the theorem can be found in Larry Wasserman's notes (Wasserman, n.d.).

Theorem 3.1 (Hammersley-Clifford). *A strictly positive probability density $f(x) > 0$ for all $x \in X$ (or strictly positive mass function $p(x) > 0$ in the discrete case) factorizes according to G if and only if it satisfies the Markov properties on G.*

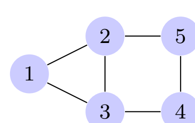
Since the distributions are strictly positive in the Hammersley-Clifford theorem, there is equivalence between the pairwise, local, and global Markov properties. The Markov properties in the theorem give an implicit characterization of the distribution through conditional independence statements, while the factorization gives a parametric characterization. Proposition 13.2.5 in (Sullivant 2018) describes factorization for undirected models in the discrete case:

Proposition 3.1. A discrete undirected graphical model associated to G consists of all distributions $p \in \Delta_{r-1}$, for $r = \prod_{i=1}^m r_i$ where $X_i \in [r_i]$ for $i \in [m]$, and $\mathcal{X} = [r_1] \times \cdots \times [r_m]$, such that

$$p_{i_1,\dots,i_m} = \frac{1}{Z(\theta)} \prod_{C\in\mathcal{C}(G)} \theta^{(C)}_{i_C}; \quad Z(\theta) = \sum_{(i_1,\dots,i_m)\in\mathcal{X}} \prod_{C\in\mathcal{C}(G)} \theta^{(C)}_{i_C}$$

where $\theta = \left(\theta^{(C)}_{i_C}\right)_{C\in\mathcal{C}(G)}$ is a vector of parameters, $i_C = (i_j)_{j\in C}$, and $Z(\theta)$ is a normalizing constant.

Example 3.3. For the undirected model G:



assume X_1, X_2, X_3, X_4, and X_5 are binary random variables. The maximal cliques of G are

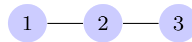
$$\mathcal{C}(G) = \{\{1,2,3\}, \{2,5\}, \{3,4\}, \{4,5\}\}$$

Then $p \in \Delta_{2^5-1}$ such that

$$p_{i_1,i_2,i_3,i_4,i_5} = \frac{1}{Z(\theta)}\theta^{(123)}_{i_1,i_2,i_3}\theta^{(25)}_{i_2,i_5}\theta^{(34)}_{i_3,i_4}\theta^{(45)}_{i_4,i_5}$$

The toric vanishing ideal of the graphical model in the previous example can be computed using the techniques described in Ch. 6 of (Sullivant 2018). This is clearly illustrated in the smaller example:

Example 3.4. The undirected model for binary X_1, X_2, X_3:



has maximal cliques $\{1, 2\}$ and $\{2, 3\}$, giving the factorization:

$$p_{ijk} = \theta_{ij}^{(12)} \theta_{jk}^{(23)}$$

The exponential family corresponding to this factorization is $p_\theta(x) = \exp(\langle \eta, Ax \rangle)$ with configuration matrix:

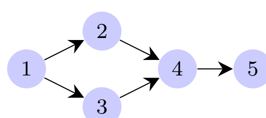
$$A = \begin{pmatrix} 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \end{pmatrix}$$

where $\theta_{i_C} = \exp(\eta_{i_C})$ for each $C \in \mathcal{C}(G)$. The toric ideal I_A is the toric ideal for this graphical model G.

3.3 Directed Parameterizations

Parameterizations of DAGs can be defined in terms of the local conditional independence statements: $X_v \perp\!\!\!\perp X_{\text{nd}(v) \setminus \text{pa}(v)} \mid X_{\text{pa}(v)}$. Applying the chain rule of probability, and ordering the vertices according to their parents (a *topological ordering*) gives the factorization of a DAG.

Example 3.5. Consider the DAG below for random variables X_1, X_2, X_3, X_4, X_5 where variables are labeled in topological order:



The chain rule factorization for the pdf $f(x)$ is

$$f(x_1, x_2, x_3, x_4, x_5) = f_1(x_1) f_2(x_2 \mid x_1) f_3(x_3 \mid x_1, x_2) f_4(x_4 \mid x_1, x_2, x_3) f_5(x_5 \mid x_1, x_2, x_3, x_4)$$

Applying the local independence statements gives the factorization:

$$f(x_1, x_2, x_3, x_4, x_5) = f_1(x_1) f_2(x_2 \mid x_1) f_3(x_3 \mid x_1) f_4(x_4 \mid x_2, x_3) f_5(x_5 \mid x_4)$$

Notice that this is the same as:

$$f(x_1, x_2, x_3, x_4, x_5) = f_1(x_1 \mid x_{\text{pa}(1)}) f_2(x_2 \mid x_{\text{pa}(2)}) f_3(x_3 \mid x_{\text{pa}(3)}) f_4(x_4 \mid x_{\text{pa}(4)}) f_5(x_5 \mid x_{\text{pa}(5)})$$

In fact, this factorization holds for all DAGs, regardless of their order, as long as the directed local Markov property holds. Definition 13.2.9 and theorem 13.2.10 from (Sullivant 2018) gives this factorization in general:

Theorem 3.2. *The parametric directed graphical model associated to the DAG $G = (V, E)$ consists of all probability densities that factorize as the product of their conditionals (recursive factorization property):*

$$f(x) = \prod_{v \in V} f_v(x_v \mid x_{\text{pa}(v)})$$

A probability density satisfies this property for an associated DAG G if and only if it satisfies the directed local Markov property associated to G.

In the discrete case this factorization can be written as the parametrization:

$$p_{i_1,\ldots,i_m} = \prod_{j=1}^{m} \theta^{(j)}(i_j \mid i_{\mathrm{pa}(j)})$$

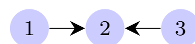
where $\theta^{(j)}(i_j \mid i_{\mathrm{pa}(j)})$ are the conditional probabilities corresponding to $X_j \in [r_j]$ such that

$$\sum_{k=1}^{r_j} \theta^{(j)}(k \mid i_{\mathrm{pa}(j)}) = 1$$

for each $i_{\mathrm{pa}(j)} = (i_l)_{l \in \mathrm{pa}(j)}$. In other words:

$$\theta^{(j)}(i_j \mid i_{\mathrm{pa}(j)}) = \mathrm{P}(X_j = i_j \mid X_{\mathrm{pa}(j)} = i_{\mathrm{pa}(j)})$$

Example 3.6. For binary X_1, X_2, X_3 modeled by the DAG



The factorization is

$$p_{ijk} = \theta^{(1)}(i)\theta^{(2)}(j)\theta^{(3)}(k \mid i, j)$$

With constraints on the parameters:

$$\begin{aligned} \theta^{(1)}(0) + \theta^{(1)}(1) &= 1, & \theta^{(2)}(0) + \theta^{(2)}(1) &= 1, \\ \theta^{(3)}(0 \mid 0,0) + \theta^{(3)}(1 \mid 0,0) &= 1, & \theta^{(3)}(0 \mid 0,1) + \theta^{(3)}(1 \mid 0,1) &= 1, \\ \theta^{(3)}(0 \mid 1,0) + \theta^{(3)}(1 \mid 1,0) &= 1, & \theta^{(3)}(0 \mid 1,1) + \theta^{(3)}(1 \mid 1,1) &= 1 \end{aligned}$$

Given these constraints, the parameters can be re-written:

$$\begin{array}{lclllcl} a & = & \theta^{(1)}(0) & \text{such that} & 1-a & = & \theta^{(1)}(1) \\ b & = & \theta^{(2)}(0) & \text{such that} & 1-b & = & \theta^{(2)}(1) \\ c_{ij} & = & \theta^{(3)}(0 \mid i,j) & \text{such that} & 1-c_{ij} & = & \theta^{(3)}(1 \mid i,j) \end{array}$$

There is an additional constraint on the distribution:

$$\sum_{ijk} \theta^{(1)}(i)\theta^{(2)}(j)\theta^{(3)}(k \mid i,j) = \sum_{ijk} p_{ijk} = 1$$

To account for this constraint, a homogenization parameter t can be included in the parameterization to get:

$$\begin{array}{lcllcl} p_{000} & = & tabc_{00}, & p_{100} & = & t(1-a)bc_{10}, \\ p_{001} & = & tab(1-c_{00}), & p_{101} & = & t(1-a)b(1-c_{10}), \\ p_{010} & = & ta(1-b)c_{01}, & p_{110} & = & t(1-a)(1-b)c_{11}, \\ p_{011} & = & t(1-b)(1-c_{01}), & p_{111} & = & t(1-a)(1-b)(1-c_{11}) \end{array}$$

The vanishing ideal of the parametrization is equal to the conditional independence ideal of the directed global Markov independence statements. See the Macaulay2 code below.

```
cat > tmp.m2 << EOF
loadPackage "GraphicalModels";

-- Define the graph in our example
G = digraph { {1,{3}}, {2,{3}} };

-- The ring of binary variables
R = markovRing(2,2,2);

-- Ideal generated by the global CI statements (1,2|3)
I = conditionalIndependenceIdeal(R, globalMarkov G);
print "\nConditional Independence Ideal generators:"
geni = toString (- gens I);
<< wrap(60,substring(9, length(geni) - 11, geni)) << endl;

-- Compare this ideal to the vanishing ideal for the parameterization
-- Ring for the parameterization
-- (the subscripts of cij changed: 0->1, 1->2)
S = QQ[t,a,b,c11,c12,c21,c22];

-- Parameterization map
f = map(S,R, {
  t*a*b*c11,         t*a*b*(1-c11),
  t*a*(1-b)*c12,     t*a*(1-b)*(1-c12),
  t*(1-a)*b*c21,     t*(1-a)*b*(1-c21),
  t*(1-a)*(1-b)*c22, t*(1-a)*(1-b)*(1-c22)
});


-- The vanishing ideal for the parameterization
print"\nVanishing Ideal generators:"
J = kernel f;
genj = toString gens J;
<< wrap(60,substring(9, length(genj) - 11, genj)) << endl;

-- It's easier to see this with a Mobius mapping:
P = markovRing(2,2,2, VariableName => "q");
m = map(R,P, {
  sum gens R,                -- p_+++
  sum (gens R)_{0,2,4,6},    -- p_++1
  sum (gens R)_{0,1,4,5},    -- p_+1+
  sum (gens R)_{0,1,2,3},    -- p_1++
  sum (gens R)_{0,4},        -- p_+11
  sum (gens R)_{0,2},        -- p_1+1
  sum (gens R)_{0,1},        -- p_11+
  p_(1,1,1)                  -- p_111
});

print"\nCI Ideal generators (Mobius Coords):";
<<  (inverse m) (-gens I) << endl;
print"\nVanishing Ideal generators (Mobius Coords):"
<<  (inverse m) (gens J) << endl;
EOF
```

```
M2 --script tmp.m2
rm tmp.m2
```

```
##
## Conditional Independence Ideal generators:
## p_(1,2,1)*p_(2,1,1)+p_(1,2,2)*p_(2,1,1)+p_(1,2,1)*p_(2,1,2)+
## p_(1,2,2)*p_(2,1,2)-p_(1,1,1)*p_(2,2,1)-p_(1,1,2)*p_(2,2,1)-
## p_(1,1,1)*p_(2,2,2)-p_(1,1,2)*p_(2,2,2)
##
## Vanishing Ideal generators:
## p_(1,2,1)*p_(2,1,1)+p_(1,2,2)*p_(2,1,1)+p_(1,2,1)*p_(2,1,2)+
## p_(1,2,2)*p_(2,1,2)-p_(1,1,1)*p_(2,2,1)-p_(1,1,2)*p_(2,2,1)-
## p_(1,1,1)*p_(2,2,2)-p_(1,1,2)*p_(2,2,2)
##
## CI Ideal generators (Mobius Coords):
## | q_(1,2,1)q_(1,2,2)-q_(1,1,1)q_(2,2,1) |
##
## Vanishing Ideal generators (Mobius Coords):
## | q_(1,2,1)q_(1,2,2)-q_(1,1,1)q_(2,2,1) |
```

References

Koller, Daphne, and Nir Friedman. 2009. *Probabilistic Graphical Models: Principles and Techniques - Adaptive Computation and Machine Learning*. The MIT Press.

Lauritzen, Steffen. 2011. "Graphical Models and Inference Mt11." https://www.stats.ox.ac.uk/~steffen/teaching/gm11/overview.htm.

Sullivant, S. 2018. *Algebraic Statistics*. Graduate Studies in Mathematics. American Mathematical Society. https://books.google.com/books?id=19V7DwAAQBAJ.

Wasserman, Larry. n.d. "36-708 Statistical Methods for Machine Learning." https://www.stat.cmu.edu/~larry/=sml/.