Conditional independence in DGM

Basic 3 nodes graphs

1) Markov chain

\[ p(x, y, z) = p(x)p(z|x)p(y|z) \]

(exercise) \[ \Rightarrow p(x, y|z) = p(x|z)p(y|z) \]

2) Latent cause

\( Z \) is not observed

3) Explaining away / competing effect

\( Z \) is not observed

Non-monotonic property of conditioning:

\[ p(\text{Alien}) \quad \text{tall} \]
\[ p(\text{Alien} | \text{tall}) > p(\text{Alien}) \]
\[ p(\text{Alien} | \text{tall, brown}) < p(\text{Alien} | \text{tall}) \]

More cond. indep. statements in a DGM

let \( \text{nd}(i) \equiv \{ j : i \notin \pi_j \} \) be the non-descendants of \( i \)

\[ p \in \mathcal{S}(\mathcal{G}) \Leftrightarrow X_i \perp \!\!\!\!\perp X_{\text{nd}(i)} \mid X_P \cup V \]
(by decomposition \[\Rightarrow \Xi_{i \forall j \mid X_{\pi_r^i} \notin \text{ND}(i)}\])

**proof:**

\[\Rightarrow\) **Key point:** let \(i\) be fixed; then \(\exists\) a top order of \(\sigma = \text{ND}(i)\) are exactly before \(i\):

\[\text{ie } (\text{ND}(i), \preceq, \text{descendants}(i))\]

\[p(x_i, x_{\text{ND}(i)}) = p(x_i | x_{\text{ND}(i)}) \prod_{j \in \text{ND}(i)} p(x_j | x_{\text{ND}(i)})\]

\[p(x_i | x_{\text{ND}(i)}) = p(x_i | x_{\text{ND}(i)}) = p(x_i | x_{\text{ND}(i)}) \prod_{j \in \text{ND}(i)} p(x_j | x_{\text{ND}(i)}) = p(x_i | x_{\text{ND}(i)})\]

\[\prod_{j \in \text{ND}(i)} p(x_j | x_{\text{ND}(i)})\]

\[\prod_{j \in \text{ND}(i)} p(x_j | x_{\text{ND}(i)})\]

\[\Rightarrow p \in \mathcal{S}(\sigma)\]

\[\Leftarrow\)**(suppose \(p\) satisfies all these cond. indep. statements)**

let \(i: j\) be \((\mathcal{W}, \mathcal{O}, \mathcal{S})\) a top sort of \(\sigma\)

then have \(1: i-1 \leq \text{ND}(i) \forall i \) [by top sort property]

\[\Rightarrow X_{i:1} X_{i-1} \mid X_{\pi_r^i} \text{ by decomposition}\]

\[p(x_i) = \prod_{j \in \text{ND}(i)} p(x_i | x_{i-1}) \text{ (by chain rule)}\]

\[= \prod_{j \in \text{ND}(i)} p(x_i | x_{\pi_r^i}) \text{ (by cond. indep.)}\]

\[\Rightarrow p \in \mathcal{S}(\sigma)\]

<table>
<thead>
<tr>
<th>cond. indep statements?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chain from (a) to (b): just any (undirected) path in graph from (a) to (b)</td>
</tr>
</tbody>
</table>

**\(d\)-separation**:

<table>
<thead>
<tr>
<th>Def: set (A) &amp; (B) are said to be (d)-separated by (C) in (\sigma)</th>
</tr>
</thead>
<tbody>
<tr>
<td>if all chains from (a \in A) to (b \in B) are &quot;blocked&quot; given (C)</td>
</tr>
<tr>
<td>where a chain from (a) to (b) is &quot;blocked&quot; given (C) at a node of a subpath of chain</td>
</tr>
</tbody>
</table>
where a chain from a to b is "blocked" given C at a node d.

if a) either d ∈ C and (vi−1, d, vi) is not a v-structure
   (e.g., vi−1, d, vi+1) or
   vi−1, d, vi+1

b) d ∈ C and (vi−1, d, vi) is a v-structure
   and no descendant of d is in C

prop:

\[ p \in \mathcal{F}(G) \iff X_a \perp \!\!\!\!\!\!\perp X_b \mid X_c \quad \forall A, B, C \subseteq V \]

or. At B are d-separated given C

"Bayes ball" alg

"intuitive" alg to check d-separation
rules balls/chains being blocked

1) in d-separation

a) 1)

b) 3)

or Vi−1, Vi

See Alg. 3.1
in K. F. book

more properties of DGM

inclusion: ECE \ ==> \ \mathcal{F}(G) \subseteq \mathcal{F}(G')
reversal. if $G$ is a directed tree (or a forest)
let $E'$ be another directed tree (with same undirected edges) by
choosing a different root

$\Rightarrow$ $s(G) = s(G')$

replacing: any directed tree for an undirected tree gives the same DGFM

Marginalization:
- marginalizing a leaf node $n$ gives a smaller DGFM
  let $S = \mathbb{E}_q$ is a dist. on $x_{i:n-1}$ s.t. $q(x_{i:n-1}) = \frac{p(x_{i:n})}{x_n}$
  for some $p \in s(G)$
  then $S = s(G')$ where $G'$ is $G$ with leaf $n$ plucked (removed)
- not true for all marginalizations
  e.g. get a set of dist.
  $\neq$ to any $s(G')$ for any $G'$
  i.e. marginalization is not a "closed" operation on DGFM's

\[
p(x_{\text{new}} | x_{i:n}) = \sum_{x_n} p(x_{\text{new}} | x_n)p(x_{i:n})
\]
undirected GM (UGM) (aka, Markov random field or Markov network)

let \( G = (V, E) \) be an undirected graph

let \( \mathcal{C} \) be the set of cliques of \( G \)

\[ \text{clique} = \text{fully connected set of nodes} \]

\( \text{if } C \in \mathcal{C} \Rightarrow \forall i, j \in C, \ x_i \neq x_j \)

UGM associated with \( G \)

\[ Z(G) \triangleq \sum_{\mathbf{x}} \mathbf{p} \text{ is a dot over } X_V \]

\[ p(x_V) = \frac{1}{Z} \prod_{C \in \mathcal{E}} \psi_C(x_C) \]

for some "potentials" \( \psi_C : \prod_{x_C} \rightarrow \mathbb{R}^+ \)

and \( Z \triangleq \sum_{\mathbf{x}} \left( \prod_{C \in \mathcal{E}} \psi_C(x_C) \right) \)

"partition function"

Notes:

- unlike a DGM, \( \psi_C(x_C) \) is not directly related to \( p(x_C) \)

- can rescale any potential without changing joint \( p \) i.e. \( \psi'_C(x_C) \triangleq \psi_C(x_C) \cdot A \) (unlike in DGM, where \( f_i \)'s were uniquely defined \( p \) for some \( A > 0 \)

- it is sufficient to consider \( \mathcal{C} \) with the set of maximal cliques

  \[ \text{e.g., } C' \subseteq C \]

  redefine \( \psi'_C(x_C) \triangleq \psi^{old}_C(x_C) \cdot \psi^{old}_{C'}(x_{C'}) \)

  (if \( C' \) belongs to more than one \( C \)
  just pick one)

- we'll see later sometimes convenient to consider "non-maximization" at trees

\[ \text{let } \mathcal{C}_1, \mathcal{C}_2 \text{ be maximal cliques} \]

\[ \mathcal{C}_1 \subseteq \mathcal{C}_2 \]

\[ \mathcal{C}_1 \cup \mathcal{C}_2 \text{ is not maximal} \]

\[ \mathcal{C}_1, \mathcal{C}_2 \text{ is maximal clique} \]
properties of UGM:

- as before \( E \subseteq E' \Rightarrow \mathcal{S}(E) \subseteq \mathcal{S}(E') \)
  \[ E = \emptyset \Rightarrow \mathcal{S}(E) = \text{set of fully factorised dist} \]
  \[ E = \text{all pars} \Rightarrow \mathcal{S}(E) = \text{all distributions on } x_v \]
  (i.e. each \( x_v \) is a clique in \( G \))

- if \( \Psi(x_v) > 0 \) \( \forall x_v \), \( \forall c \)
  then can write \( p(x_v) = \exp \left( \sum \log \Psi(x_v) - \log Z \right) \)

\[
\Theta_c(x_v) = \log \Psi_c(x_v)
\]

\[
\text{dim } T_c = 1 \quad \text{or} \quad 0
\]

\[
\mathbf{T}_c(x_v) = \begin{cases} 
0 & \text{if } x_v = 1 \\
1 & \text{if } x_v = 0 
\end{cases}
\]

- physics link: negative energy field

- e.g. Ising model in physics

- node potential \( E_i = \log \Psi_i(x_i=1) \)
- edge potential \( E_{ij} = \log \Psi_{ij} (x_i=1, x_j=1) \)

other's social network modelling