**Conditional independence in DGM**

1. **Markov chain**
   
   \[
   p(x, y, z) = p(y|x)p(z|x)p(x) \quad \text{but} \quad X \perp Y | Z
   \]
   
   (exercise) \Rightarrow \ p(x,y|z) = p(z|x)p(y|x)

2. **Latent cause (hidden variable)**
   
   \[
   \begin{aligned}
   &X \quad \text{age} \quad \\
   &\text{smoke} \times \quad \text{signs} \times \quad \text{gray hair}
   \end{aligned}
   \]

3. **Explaining away/compacting effect**
   
   \[
   \begin{aligned}
   &X \quad \text{watch is broken} \quad \\
   &\text{watch is broken} \quad \text{in late} \quad \text{but} \quad Y \perp Z
   \end{aligned}
   \]

   (for some p)

**Non-monotonic property of conditioning**

- \( p(\text{alien}) \) tiny
- \( p(\text{alien}) \text{ late} \geq p(\text{alien}) \)
- \( p(\text{alien} \mid \text{late}, \text{watch is broken}) > p(\text{alien} \mid \text{late}) \)

**More cond. indep. statements in DGM**

let \( \text{nd}(i) = \{ j \mid j \neq i \text{ and no path from } i \text{ to } j \} \)

"non-descendant of i"

**Proof**

\[ p \in \mathcal{S}(G) \iff X_i \perp X_{\text{nd}(i)} \mid X_j, \forall j \in \text{nd}(i) \]

(by decomposition: \Rightarrow \ X_i \perp X_{\text{nd}(i)} \mid X_j, \forall j \in \text{nd}(i))

**Proof**

\[ \Rightarrow \text{ key point: let } i \text{ be fixed, then } G \text{ a bp. ordering of } G \text{ satisfies } \]

\( \text{nd}(i) \text{ are exactly before } i \text{ i.e. (nd}(i), i \text{, descendant}(i)) \)

\[ \Rightarrow p(X_i, X_{\text{nd}(i)}) = p(X_i \mid X_{\text{nd}(i)}) \prod_{j \in \text{nd}(i)} p(X_j \mid X_{\text{nd}(i)}) \]

Please add more
\[
p(x_1; x_{Ind.0}) = p(x_1, x_{Ind.1}) = p(x_1; x_{Ind.1}) \prod_{j \in Ind.1} p(x_j | x_{Ind.1}) = p(x_1; x_{Ind.1})
\]

\[
\text{(suppose \( p \) satisfies all these cond. indep. statements)}
\]

\[
\text{let } 1:n \text{ be (undirected) a top. xnt. of } G
\]

\[
\Rightarrow X_i \perp \!\!\!\!\!\!\perp X_{i:t-1}, x_{i:t}, \text{ by decomposition}
\]

\[
p(x_v) = \prod_{i=1}^{n} p(x_i | x_{i:t-1}) \text{ (by chain rule)}
\]

\[
= \prod_{i=1}^{n} p(x_i; x_{i:t}) \text{ (by cond. indep.)}
\]

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

\[\text{other cond. indep. statements?}\]

\[\text{chain from a to b: just any (undirected) path in graph from a to b}\]

\[\text{\( d \)-separation: }
\]

\[\text{def: set } A \& B \text{ are said to be \( d \)-separated by } C \text{ (in } G) \text{ conditioned on set } \{A,B\}
\]

\[\text{iff all chains from a to b are "blocked" given } C
\]

\[\text{where a chain from } a \text{ to } b \text{ is "blocked" given } C \text{ at node } d
\]

\[\text{if a) either } d \in C \text{ and } (v_{i-1}, d, v_i) \text{ is not a } v\text{-structure}
\]

\[\text{b) } d \notin C \text{ and } (v_{i-1}, d, v_i) \text{ is a } v\text{-structure and no descendant of } d \text{ is } v_i \in C
\]

\[\text{prep: } p \in S(\mathcal{G}) \iff X_A \perp \!\!\!\!\!\!\perp X_B \perp X_C \text{ \& } A,B,C \subseteq V
\]

\[\text{iff A \& B are d-separated by } C
\]

\[\text{\"Bayes ball" alg.}
\]

\[\text{\"intuitive" alg. to check } d\text{-separation}
\]

\[\text{rules balls/chains being blocked}
\]
more properties of DGM

- Inclusion: \( E \subseteq E' \Rightarrow s(E) \subseteq s(E') \)
- Reversal: if \( G \) is a directed tree (or a forest) \( \Rightarrow \) there is no \( v \)-structure

- Let \( E' \) be another directed tree (with same undirected edges) by choosing a different root \( \Rightarrow s(E) = s(E') \)
- Reversing: all directed trees from an undirected tree give same DGM

- Marginalization:
  * Marginalizing a leaf node \( n \) gives us a smaller DGM
    \[
    s(S) = \sum_{n} \prod_{i \neq n} q(x_i, n) = \prod_{i \neq n} p(x_i, n) \forall n \in S
    \]
    \[
    s(S) = s((G) \setminus n) \text{ where } G' \text{ is } G \text{ with leaf } n \text{ plucked (removed)}
    \]
* Not true for all marginalizations. Ex. "marginalize out this node"

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

Let \( \mathcal{C} \) be the set of cliques of \( G \). A clique is a fully connected set of nodes.

Undirected GM (UGM) (aka. as Markov random field or Markov network)

\[ f(G) \triangleq \sum_{\mathcal{C}} \text{p is a dist over } x \]

for some potentials \( \psi_c : \{x\} \to \mathbb{R}_+ \)

\[ \mathcal{Z} \triangleq \sum_{\mathcal{C}} \text{p}(x_\mathcal{C}) \]

\[ \text{Normalization constant} \]

\text{"partition function"}

Notes:

- Unlike a DGM, \( \psi_c(x) \) is not directly related to \( p(x) \).

- Can rescale any potential without changing the joint.

- It is sufficient to consider \( \psi_{\text{max}} \), the set of maximal cliques.

\[ \mathcal{C}_{\text{max}} \subseteq \mathcal{C} \]

Redefine \( \psi_c^{\text{max}}(x_c) \triangleq \psi_c(x_c) \cdot \psi_c^{\text{d}}(x_c) \)

(Just choose one)

\[ \text{e.g. } C^2 \subseteq C \]
properties of UGM:

- as before  \( E \subseteq E' \Rightarrow \mathcal{Z}(E) \subseteq \mathcal{Z}(E') \)
  \[ E = \emptyset \Rightarrow \mathcal{Z}(E) = \text{set } \text{fully synchronized dist.} \]
  \[ E = \text{all pairs } \Rightarrow \mathcal{Z}(E) = \text{all dist. on } x \]
  (i.e., \( V \) is
  just a big clique in \( G \))

- if \( \psi_c(x_c) > 0 \) then
  \[ \Psi(x) = : \exp \left( \sum_{x_c} \log \psi_c(x_c) - \log z \right) \]
  \[ \alpha(x) = |x_c| \]
  \[ \phi(x) = \text{log } \psi_c(x_c) \]
  \[ \mathcal{T}(x) = \text{log } \psi_c(x_c) \]
  \[ \mathcal{Z} = \sum_{x_c} \mathcal{T}(x) \psi_c(x_c) \]
  \[ <\Theta_c, \mathcal{T}(x_c)> \text{ where } \Theta_c(x) = \left( \begin{array}{c} 1 \\ 0 \end{array} \right) \]
  \[ \Theta_c(x) = \text{log } \psi_c(x_c) \]

  physics link \( \rightarrow \) negative energy

  e.g. Ising model in physics

  \( x \in \{0,1\} \)

  node potential  \( E = -\text{log } \psi_i(x_i) \)

  edge potential  \( E_{ij} = \text{log } \psi_{ij}(x_i \neq x_j) \)

  other examples: social network