today: finish DGM
start UGM

cond. indep. statements in DGM:

\[ \text{let } \text{nd}(i) \triangleq \{ j : \text{no path } i \rightarrow j \} \]

"non-descendants" of \( i \)

prop.:

\[ p \in \mathcal{B}(\mathcal{G}) \iff \forall i \exists j \in \text{nd}(i) \subseteq \text{par}(i) \]

(by decomposition \( \Rightarrow \) \( \forall i \exists j \in \text{nd}(i) \subseteq \text{par}(i) \))

proof:

\[ \Rightarrow \]

key point: let \( i \) be fixed,
then \( i \) a top ordering s.t. \( \text{nd}(i) \) are exactly before \( i \)

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

pick all these:

\[ p(\mathbf{x}_i, \mathbf{x}_{\text{nd}(i)}) = p(x_i | x_{\text{nd}(i)}) \prod_{j \in \text{nd}(i)} p(z_j | x_{\mathbf{z}_j}) \]
\[
p(x_i | x_{\text{nd}(i)}) = \frac{p(x_i, x_{\text{nd}(i)})}{p(x_{\text{nd}(i)})} = p(x_i | x_{\text{nd}(j)}) \prod_{j \in \text{nd}(i)} \frac{p(x_i | x_{\pi_j})}{p(x_i)} = p(x_i | x_{\pi_j})
\]

\[
\rightarrow \quad \prod_{j \in \text{nd}(i)} \frac{p(x_i | x_{\pi_j})}{p(x_i)} = 1 \\
\text{(because } j \text{ is nd}(i) \Rightarrow i \neq \pi_j \text{)}
\]

\[\] (suppose \( p \) satisfies cond. indep. statements)

\[\text{Let } \mathfrak{x} \text{ be a fp sort}
\]

\[
\Rightarrow \quad X_i \perp X_{i:i-1} | X_{\pi_i} \quad \text{(by decomposition)}
\]

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

\[
= \prod_{i=1}^n p(x_i | x_{\pi_i}) \quad \text{(by C.I.)}
\]

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

\[\]

\[\] (other C.I. statements?)

\[\] (chain ... just any (undirected) path in graph from \( a \) to \( b \))
Chain from a to b: just any (undirected) path in graph from a to b

\[ a \text{-} \text{separation} \]: intuition, i.e., do not want this situation

\[ \text{Def:} \] sets A \& B are said to be \textbf{d-separated} by C iff all chains from a \& A to be \& B are "blocked" given C

where a chain from a \& b is "blocked" at node d if it is a subpath of chain

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

\[ \text{eg: } v_{i-1} \xrightarrow{d} v_i \xrightarrow{v_{i+1}} \]

b) d \( \in \) C and (\( v_{i-1}, d, v_{i+1} \)) is a v-structure

\[ v_{i-1} \xrightarrow{d} v_i \xrightarrow{v_{i+1}} \text{ and no descendant of } d \text{ is in } C \]

\[ \text{Prop: } \frac{P \in S(6)}{\text{iff } X_{\text{all}} \times B \mid X_{C} \text{ i.e., no conditioning}} \]
Theorem 1.1: \[ \text{If } p \in \mathcal{F}(6) \iff \forall X \in \mathcal{X} \setminus B \mid X \subseteq A \wedge B \text{ are d-separated by } C \]

"Bayes-ball" alg.: "intuitive" alg. to check d-separation rules for balls/chains being blocked.

1) [Diagram showing d-separation in a graph with arrows indicating paths being blocked]

2) [Diagram showing another d-separation scenario]

3) V-structure

see Alg. 2.1
Properties of DBM:

- Inclusion: \( E \subseteq E' \) \( \Rightarrow \) \( S(G) \subseteq S(G') \)
  
  (adding edges increase \# of distributions)

- Reversal: if \( G \) is a directed tree (or forest), \( \Rightarrow \left| \mathcal{E}_r \right| \leq 1 \); i.e., there is no \( v \)-structure.

Then let \( E' \) be another directed tree by choosing a different root (reverse some edges, while keeping a directed tree)

\[ \Rightarrow S(G) = S(G') \]

\[ \Rightarrow [G' = (V; E')] \]

\( \odot \) this is why directing edges are not causal
\[ L(G = (V, E)) \]

\[ \text{e.g., } p(x, y) = p(z_1, y)p(y) = p(y|x)p(x) \]

- this is why direction of edges are not causal

---

Rephrasing: all directed trees from a undirected tree give same DGM

Marginalizing:

- Marginalizing a leaf node \( n \) gives us a smaller DGM.

  \[ S = \sum_{x} q \text{ is dist to } x_{:n-1} \]

  \[ q(x_{:n-1}) = p(x_{:n-1}) \text{ for } p \in S(G) \]

  Then \( S = S(G') \) where \( G' \) is \( G \) with leaf \( n \) removed.

- Not true for all marginalization.

  E.G.

  \[ \text{get a set of distributions } \neq \{ \text{any } S(G') \} \text{ for } G' \]
i.e. marginalization is not a "closed operation" on DGM

\[ \text{undirected GM (UGM)} \quad \text{(aka. Markov random field or Markov network)} \]

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

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

- A clique is a fully connected set of nodes
  \[ (\text{i.e. } C \in \mathcal{C} \Rightarrow \forall i, j \in C, \ i \neq j \Rightarrow \{i, j\} \in E) \]

- UGM associated with \( G \):
  \[ \mathcal{Z}(G) \triangleq \sum_{x} \rho \quad \text{s.t.} \quad \rho(x) = \frac{1}{\mathcal{Z}(C)} \prod_{C} \psi_{C}(x_{C}) \]

For some "potentials" \( \psi_{C} \) s.t.

\[ \psi_{C}(x_{C}) > 0 \quad \forall x_{C} \]

and \( \mathcal{Z} \triangleq \sum_{x} \prod_{C} \psi_{C}(x_{C}) \) "partition function"
and \( Z = \sum_{C} (\prod_{c} \psi_c(x_c)) \) "partition function"

**Notes:**
- Unlike in a DGM (where we could think of \( C \) to be \( (i, m_i) \), \( \psi_c(x_c) = p(x_i | z_{i}) \)), \( \psi_c(x_c) \) is not directly related to \( p(x_c) \).
- Can multiply any \( \psi_c(\cdot) \) by a constant without changing \( p_c \).
- It is sufficient to consider only \( C_{\max} \), the set of maximal cliques.

\[ C' \subseteq C \]

Redeclare \( \psi_{c'}(x_c) = \psi_{c,\text{old}}(x_c) \psi_{c',\text{new}}(x_c) \)

**Properties:**
- As before, \( E \subseteq E' \Rightarrow S(E) \leq S(E') \)

\[ E = \emptyset \quad \Rightarrow \quad S(E) = \text{fully factorised dist.} \]

\[ E = \text{all pairs} \quad (\text{i.e., } G \text{ is one big clique}) \quad \Rightarrow \quad \text{all distributions} \]
• If \( \Phi_c(x_c) > 0 \) for \( x \neq x_c \),
  can write \( p(x) = \exp \left( \frac{1}{kT} \log \Phi_c(x_c) - \log Z \right) \).

  **Physics link:** negative energy function

  *E.g., Ising model in physics:*

  \[ x_i \in \{0, 1\} \]

  node potentials \( \rightarrow E_i \)

  edge potentials \( \rightarrow E_{ij} \)

  *Think also social network modeling*

  **Conditional independence for UGM:**

  **Def:** we say that \( p \) satisfies the **global Markov property** *(with respect to undirected graph \( G \))*

  iff \( \forall A, B, S \subseteq V \text{ s.t. } S \text{ separates } A \text{ from } B \text{ in } G \)
Prop: \( p \in S(\mathcal{G}) \Rightarrow p \) satisfies the global Markov property

Proof: WLOG, we assume \( A \cup \text{BUS} = V \)

[why? if not, let \( \tilde{A} = A \cup \exists a \in V: a \text{ and } \tilde{A} \text{ are not separated by } S \) \( \tilde{B} = V \setminus (S \cup \tilde{A}) \)]

then if have \( X_A \perp \perp X_B \mid X_S \Rightarrow X_A \perp \perp X_B \mid X_S \) \[ \text{by decomposition} \]

let \( C \in \mathcal{G} \)

cannot have \( C \cap A \neq \emptyset \) and \( C \cap B \neq \emptyset \)

thus \( q(x) = \frac{1}{\sum_{C \in \mathcal{G}} \psi_C(x)} \left( \sum_{C' \in \mathcal{G}} \psi_{C'}(x') \right) = f(x_{\text{AUS}}) g(x_{\text{BUS}}) \)

\( \Rightarrow C \subseteq \text{BUS} \)
\[ p(x_A | x_S) \propto p(x_A, x_S) = \frac{f(x_A, x_S)}{f(x_B)} \geq \frac{f(x_A, x_S)}{f(x_B)} \]

\[ \Rightarrow p(x_A | x_S) = \frac{f(x_A, x_S)}{\sum_{x_A} f(x_A, x_S)} \]

Similarly, \[ p(x_B | x_S) = \frac{g(x_B, x_S)}{\sum_{x_B} g(x_B, x_S)} \]

\[ p(x_A | x_S) p(x_B | x_S) = \frac{f(x_A, x_S) g(x_B, x_S)}{\sum_{x_A, x_B} f(x_A, x_S) g(x_B, x_S)} \geq p(x_V) = p(x_A, x_B | x_S) \]

\[ \geq \frac{\sum_{x_A, x_B} f(x_A, x_S) g(x_B, x_S)}{\sum_{x_A, x_B} f(x_A, x_S) g(x_B, x_S)} \]

\[ \text{i.e., } x_A \perp x_B | x_S \]

**Thm:** (Hammond-Clifford) \[ \text{if } p(x_V) \geq 0 \forall x_V \]

Then \[ p \in \mathcal{B}(\mathcal{G}) \iff p \text{ satisfies global Markov property} \]
Then \( \pi(E) \iff \pi \) satisfies global Markov property.

**Proof:** see ch. 16 of Mike's book; use "Möbius inversion formula" (as exclusion-inclusion principle in prob.)