today: find DGM
• graph eliminate

properties of DGM:

inclusion: \( E \subseteq E' \Rightarrow \delta(E) \subseteq \delta(E') \)

(adding edges increases # of dist.)

reversal: if \( G \) is a directed tree (or forest)


\[ \forall (1 \leq r \leq 1) \text{ i.e. there is no v-structure} \]

then let \( E' \) be another directed tree by choosing a different root (reverse same edges while keeping a directed tree)

\[ \Rightarrow \delta(E) = \delta(E') \]

\( \rho(V,E) \)

\[ \forall \text{this is why edges are not causal} \]

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

\[ p(xv) = \sum p(x;|x,v) = \sum p(x; |x,v) \]

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

marginalizing:

• marginalizing a leaf gives us a smaller DGM

\[ = \sum_{\text{marginalize}} p(x_1; n) \]
i.e. let \( S = f(p) \) is set in \( x_i^e \) s.t. \( q(x_i^e) = p(x_i^e) \) for \( p \in S(\mathcal{G}) \)

then \( S = S(\mathcal{G}') \) where \( G' \) is \( G \) with leaf \( n \) removed

- not true for all marginalization:

  - e.g.

\[
\begin{array}{c}
\text{marginalizing out this node} \\
\text{get a set of dist.} \\
\neq \text{to any } S(\mathcal{G}') \text{ for some } G'
\end{array}
\]

\[
\begin{array}{c}
\text{i.e. marginalization is not a "closed operation" on } \mathcal{GM}
\end{array}
\]

\underline{Undirected GM (UGM)} (aka Markov random field or Markov network)

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

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

- clique = fully connected set of nodes

i.e. \( C \in \mathcal{G} \iff \forall i, j \in C, (i, j) \in E \)

\[
\begin{array}{c}
\text{then the UGM associated with } G
\end{array}
\]
\[ s(G) = \sum_{\omega} p(\omega) \]  where \( p(\omega) = \frac{1}{Z_{\text{ceb}}} \psi_c(\omega) \frac{1}{2} \]

\[ Z_{\text{ceb}} = \sum_{\omega} \psi_c(\omega) \]

where \( \psi_c(\omega) \geq 0 \) for all \( \omega \) "potentials"

and

\[ q = \sum_{\omega} \psi_c(\omega) \]

"partition function"

Notes:

- Unlike in a B.V. (where could think of \( \psi \) to be \( \psi(x_i) \), \( \psi_c(\omega) = p(x_i|x_i) \))
  \( \psi_c(\omega) \) is not directly related to \( p(x_c) \)

- Can multiply any \( \psi_c(\cdot) \) by a constant without changing \( p \)

- It is sufficient to consider only \( \mathcal{E}_{\text{max}} \), the set of maximal cliques

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

\[ \mathcal{C} \text{ max: } \psi_c(x_c) = \psi_c(x_c)^{\text{max}} \]

\[ \psi_c(x_c) = \psi_c^{\text{left}}(x_c) \psi_c^{\text{right}}(x_c) \]

(we'll see later e.g. "overparameterization" of tree using both \( \psi_i(x_i) \) and \( \psi_j(x_i|x_j) \)

Properties:

- As before, \( E \subseteq E' \Rightarrow s(G) \leq s(G') \)

- \( E = \emptyset \Rightarrow \) fully factorized dist.

- \( E = \) all pairs (i.e. \( \text{one big clique} \)) \( \Rightarrow \) all distribution

\[ \sum_{\omega} e^\omega = 1 \]

\[ \Rightarrow \psi_c \text{ exponential family!} \]

\[ \text{we'll see later} \]
for \( W_c(x_0) > 0 \) \( \forall x_c \) 

(see write \( p(x) = \exp \left( \sum_{c \in \mathcal{C}} \log W_c(x_c) - \log Z \right) \) 

\[ \text{negative energy function} \]

e.g. Ising model in physics: \( x_i \in \{0,1\} \) 

node potentials \( e_i \) 
edge \( \equiv e_{ij} \)

think social network modeling

Conditional independence for GSM

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

if \( A, B, S \in V \) s.t. \( S \) separates \( A \) from \( B \) in \( G \)

then \( X_A \perp \perp X_B | X_S \)

\( A \quad S \quad B \)

\( S \) separates \( A \) from \( B \)

if all paths from \( A \) to \( B \) pass through \( S \)

**Proof:** \( \rho \in \mathcal{G}(G) \) \( \Rightarrow \rho \) satisfies the **global Markov property**

**Proof:** WLOG, we assume \( A \cup B \cup S = V \)
and disjoint

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

then if have \( XX \notin Xs \), we have \( XX \notin Xs \) by decomposition

\[ V = A \cup S \cup B \]

let \( C \subseteq \tilde{B} \)

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

thus \( p(z) = \sum_{\substack{\tilde{C} \subseteq \tilde{B} \\ \tilde{C} \cap A \neq \emptyset}} \psi_C(z) \psi_0(z) = f(x_{AUS})g(x_{BUS}) \)

\[ p(x_A|x_S) + p(x_{AUS}) = \sum_{x_B} p(x_V) = \sum_{x_B} f(x_{AUS})g(x_{BUS}) \]
\[ = f(x_{AUS}) \sum_{x_B} g(x_{BUS}) \]
\[ \Rightarrow p(x_A|x_S) = \frac{f(x_A, x_S)}{\sum_{x_B} f(x_A, x_S)} \]

\[ \text{Similarly, } p(x_B|x_S) = \frac{g(x_B, x_S)}{\sum_{x_A} g(x_B, x_S)} \]
\[
p(x_A | x_S) p(x_B | x_S) = \frac{\sum_{x_B} p(x_A, x_B | x_S)}{\sum_{x_B} p(x_B | x_S)} \quad \text{and} \quad p(x) = p(x_A, x_B | x_S)
\]

\[\text{ie. } x_A \perp x_B | x_S \]

**Theorem (Hammerley-Clifford):** If \( p(x) > 0 \) \( \forall x \)

Then \( p \in \mathcal{G} \) \( \iff \) \( p \) satisfies global Markov property

\( \text{(HC)} \)

**Proof:** See ch. 16 of Mike's book; use "Markus immersion formula"

(As exclusion-inclusion principle in proof)

**Properties:**

- **Closure with respect to marginalization:**
  
  Let \( V' = V \setminus \{n\} \), \( E' = \) edges in \( G \setminus \{n\} \)
  
  + Connect all neighbors of \( n \) in \( G \) together (new clique)

  \[ \text{is } \sum \text{ marginal on } x_{\text{rest}} \text{ for } p \in \mathcal{G}(6) \]

  \[ = \mathcal{G}(6') \]

**DGM vs UGM:**

**Def:** Markov blanket for \( i \) is the smallest set of rules \( M \)
for UGM: \( M = \mathcal{S} \cup \{ \mathcal{S} : \mathcal{S} \in \mathcal{E}_2 \} \) = set of neighbors of \( i \)

for OGM: \( M = \mathcal{P}_i \cup \text{children}(i) \cup \mathcal{T}_j \text{ children}(i) \)

Recap:

\[
\begin{array}{c|c}
\text{OGM} & \text{UGM} \\
\hline
\text{factorization} & p(x) = \prod_i p(x_i | \mathcal{P}_i) \quad \frac{1}{Z} \prod_i \varphi_i(x_i) \\
\text{cond. indep.} & \alpha - \text{separation} \quad \text{separation} \\
\text{marginalization} & \text{not closed in general} \quad \text{closed} \\
& \text{but fine for leaves} \\
\end{array}
\]

cannot exactly capture:

\[
\begin{array}{c|c}
\text{v-structure} & \text{v-structure} \\
\end{array}
\]

Moralization

let \( G \) be a DAG; when can we transform to equivalent UGM?
Def: For $G$ a DAG, we call $\overline{G}$ the **moralized graph** of $G$.

where $\overline{G}$ is an undirected graph with $V$.

and $E = \{ \overline{e} : e \in E \}$.

where $\overline{e} = \{ (i, j) : e = \{i, j\} \}$ and is undirected version of $e$.

and $\overline{E} = \{ \overline{e} : e \in E \}$.

L and $E$ are edges.

For some $i$.

"moralized" "marginalizing the parents" (??)

9. Connect all the parents of $i$ with $i$ in any clique.

\[ \overline{G} \]

eg. $G$

Let only add edges when $(e, (i, j))$.

\[ \overline{G} \]

i.e. $v$-structure.

Prop: For a DAG $G$ with no $v$-structure, then $\mathcal{L}(G) = \mathcal{L}(\overline{G})$.

But in general, can only say that $\mathcal{L}(G) \leq \mathcal{L}(\overline{G})$.

[Note that $\overline{G}$ is the minimal undirected graph s.t. $\mathcal{L}(G) \leq \mathcal{L}(\overline{G})$.]