today:

- undirected graph model
- inference

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

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

Let \( C \) be the set of cliques of \( G \) \( \text{clique} = \text{fully connected set of nodes} \)

\( \text{i.e. } C \subseteq V \quad \forall i, j \in C \quad \forall (i, j) \in E \)

\[ \text{UGM associated with } G \]

\[ Z(G) = \sum_{\phi} p : p(x_v) = \frac{1}{Z} \prod_{c \in C} \psi_c(x_c) \]

for some "potentials" \( \psi_c \), i.e., \( \psi_c(x_c) \geq 0 \ \forall x_c \)

and \( Z \triangleq \sum_{x_v} \prod_{c \in C} \psi_c(x_c) \) "partition function"

\[ Z = \sum_{x_1} \ldots \sum_{x_n} \]

Note:
- Unlike in DGM (where we could think of \( C \) to be \( (i, j) \), \( \psi_c(x_c) = p(x_i | x_j) \))

\( \psi_c(x_c) \) is not directly related to \( p(x_c) \)

- Can rescale any potential without changing \( p \) i.e., \( \psi_c^{\text{new}}(x_c) = \psi_c(x_c) \cdot c^{\frac{1}{2}} \)

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

**e.g.** \( C' \subseteq C \)

Redefine \( \psi_c^{\text{new}}(x_c) = \psi_c(x_c) \cdot \psi_c^{\text{old}}(x_c) \)

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

**\( \psi_c^{\text{old}}(x_c) \)**

\[ \psi_c(x_c) \]

\[ c^{\frac{1}{2}} \]

\[ \sum \]

\[ \text{we'll see later it is convenient to} \]

\[ \ldots \ldots \]
properties:
- As before, \( E \leq E' \Rightarrow \mathcal{S}(E) \leq \mathcal{S}(E') \)
- \( E = \emptyset \Rightarrow \mathcal{S}(E) = \text{fully connected dist.} \)
- \( E = \text{all pairs} \quad \mathcal{S}(E) = \text{all distributions} \) (i.e., \( G \) is just one big clique)
- If \( \chi_i(x_i) > 0 \) \( \forall x_i \)
  - can write \( p(z) = \exp \left( \sum_{E \in \mathcal{E}} \log \chi_i(x_i) - \log \beta \right) \)
  - physical link: negative energy function

e.g. Ising model in physics
- \( x_i \in \{0,1\} \)
  - node potential \( \rightarrow \chi_i \)
  - edge potential \( \rightarrow E_{ij} \)
  - \( E_i = \chi_i(x_i=1) \)
  - \( E_{ij} = \chi_{ij}(x_i=1, x_j=1) \)

Other example: social network modeling

**Conditional independence for UGM**

Definition: We say that \( p \) satisfies the global Markov property (with respect to an undirected graph \( G \)) if, for all \( S \subseteq V \) and \( A \subseteq V \setminus S \),

- If \( A, B \in \mathcal{S} \) and \( S \) separates \( A \) from \( B \) in \( G \)
  - then \( X_A \perp \perp X_B \mid X_S \)

\[ A \overset{S}{\perp} B \]
Prop: \( p \in \mathcal{F}(G) \Rightarrow p \) satisfies the global Markov property for \( G \)

Proof:

WLOG, we assume \( A \cup B \cup S = V \). \( S \) is a separator of \( V \) by \( S \).

[Why? If not, let \( A = A \cup \{a \} \in \mathcal{A} \); \( a \) and \( A \) are not separated by \( S \).
\( B = V \setminus (A \cup \{a \}) \).

Then if we have \( X_A \perp\!\!\!\perp X_B \mid X_S \)

\( \Rightarrow X_A \perp\!\!\!\perp X_B \mid X_S \)

by decomposition property

[Exercise: Show that if \( b \in \hat{B} \)

\( \Rightarrow b \) is separated from \( A \) by \( S \).

\( V = A \cup S \cup B \)

Let \( C \in \mathcal{E} \).

Cannot have \( C \cap A \neq \emptyset \) and \( C \cap B \neq \emptyset \).

Thus \( p(z) = \prod_{z \in \mathcal{E}_{z}} \prod_{C \in \mathcal{E}_{z}} \prod_{C \cup \mathcal{A} \cup S} \prod_{C \cup \mathcal{B} \cup S} = f(z_{AUS})g(z_{BUS}) \)

\[ p(x_A | x_S) = \frac{p(x_A, x_S)}{p(x_S)} \leq p(x_V = x_B) = f(z_{AUS})g(z_{BUS}) \]

\[ \Rightarrow f(z_{AUS}) \leq g(z_{BUS}) \]

(consistent with respect to \( x_A \))
\[ p(x_A | x_S) = \frac{f(x_A, x_S)}{f(x_A, x_S) | x_S} \]

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

\[
p(x_A | x_S) p(x_B | x_S) = \frac{f(x_A, x_S) g(x_B, x_S) \cdot p(x_S)}{f(x_A, x_S) g(x_B, x_S) \cdot p(x_S)} \quad \text{implies} \quad X_A \perp X_B | X_S
\]

**Thm.: (Hammond's-Gaifman)** If \( p(x_S) > 0 \ \forall x_S \)

then \( p \in \mathcal{G}(G) \iff p \) satisfies the global Markov prop. for \( G \)

**Proof:** See ch. 6 of Mike's book; use "Markov inversion formula" (as exclusion-inclusion principle in graph)

**Properties of GGM:**

- Closure with respect to marginalization:

  \[ \text{Let } V' = V \setminus \{n\}, \quad E' = \{e \in E | e \text{ connects } V' \setminus \{n\} \text{ to } \{n\} \} \]

  \[ \{ \text{marginal on } \sigma_{i,n-1} \text{ for } p \in \mathcal{G}(G) \}^2 = \mathcal{G}(G) \]

**DGM vs UGM**

- **Def.** Markov blanket for \( i \) (for graph \( G \)) is the smallest set of nodes \( N \) at \( x_i \perp x_i | x_i \) (in G)

  - For GGM: \( N = \mathcal{G}(G) \setminus \{i\} \)
  - For UGM: \( N = \mathcal{N}(\{i\}) \)
  - \( \mathcal{N}(\{i\}) = \{ \text{set of neighbors of } i \} \)
for UGM: \( M = \{ u \} \) - set of neighbors of \( \hat{u} \)

for DGM: \( M = \pi_{\hat{u}} \uplus \text{children}(i) \uplus \bigcup_{j \in \text{children}(i)} \pi_j \)

\( \text{recap} \)

<table>
<thead>
<tr>
<th>DGM</th>
<th>UGM</th>
</tr>
</thead>
<tbody>
<tr>
<td>( p(z) = \prod p(x_i</td>
<td>z_{\pi_i}) )</td>
</tr>
<tr>
<td>d-separation</td>
<td>separation</td>
</tr>
<tr>
<td>not closed in general</td>
<td>closed</td>
</tr>
<tr>
<td>but is fine for loops</td>
<td>(cannot all neighbors of removed node)</td>
</tr>
<tr>
<td>cannot exactly capture some familial</td>
<td>V-structure</td>
</tr>
</tbody>
</table>

**Moralization:**

Let \( G \) be a DAG; when can we transform \( G \) to an equivalent UGM

**Def:** for \( G \) a DAG, we call \( \overline{G} \) the **moralized graph** of \( G \)

where \( \overline{G} \) is an undirected graph with same \( V \)

and \( E = \left\{ i \mid j : (i, j) \in E \right\} \) undirected version of \( G \)

\( U \ni k, \xi, s \ni k \neq 0 \iff \xi > s \) for some \( i \neq \) moralization

\( (\text{connect all the parents of } \hat{u} \text{ with } i \text{ in big dope}) \)
For a DAG G with no V-structure, then $S(G) = S'(G)$.

But in general, can only say $S(G) \leq S'(G)$.

Note that $\tilde{G}$ is the minimal undirected graph so $S(G) \leq S(\tilde{G})$.

General themes in this class:

A) representation $\rightarrow$ DGM $\rightarrow$ VGM
   parameterization $\rightarrow$ exponential family

B) inference $p(Z | x)$ $\rightarrow$ today's elimination alg.
   "query" (and more)
   $\rightarrow$ today's elimination alg.
   "query" (and more)
   $\rightarrow$ sum-product/ belief propagation
   (for trees)

C) statistical estimation/ learning $\rightarrow$ MLE
   maximum entropy
   method of moments

Inference:

present alg. for VGM for simplicity and more generally

but note that sometimes more efficient...
make $DGM \rightarrow UGM$ via renormalization

i.e. $DGM : p(z) \propto \prod_i p(x_i | z_i) \propto \prod_i \psi(z_i)$

\[ C = \frac{1}{Z} \sum_{z_i} \psi(z_i) \]

\[ Z = \prod_i p(x_i | z_i) \]

\[ \frac{1}{Z} \prod_i \psi(z_i) \]

inference: want to compute

a) marginal: $p(z_F)$ for some FSV

b) conditional: $p(z_E | z_F)$

"query" "evidence"

c) for UGM: partition function $Z = \prod \psi(z_i)$

Why?

* missing data

  $p(z_{\text{unde}} | z_{\text{obs}})$

  * prediction e.g. $p(z_{\text{futur}} | z_{\text{past}})$

  "latent cause" e.g. $p(z_{\text{cause}} | z_{\text{obs}})$

* also related to inference

  \[ \text{argmax}_{z_F} p(z_F | z_E) \]

  Found in big hae (e.g. speech recognition)
* Inference is also needed during estimation (parametric setting MLE)

\[ \text{[e.g. during E-step \( p(\mathbf{z} | \mathbf{x}) \)]} \]