Recap.

- Probabilistic Graphical Models and Markov Property. ✓
- Inference Problems — efficient algorithms. ←
  - variational methods & Sampling.
- Learning graphical models.
- Extra topics.

Today: Efficient algorithms to approximately solve inference problems.

Inference Problems:

given an undirected graphical model

$$P(x) = \frac{1}{Z} \prod_{c \in \mathcal{C}} f_c(x_c)$$

$\mathcal{C}$ ← set of maximal cliques

① calculate marginal $P(x_i)$ ←

② calculate $\arg\max_x P(x)$ ←

③ calculate partition function Z

④ Sample from $P(x)$

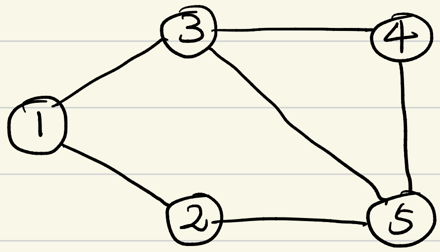
example: Y ~ observation, X ~ cause, state, $P(x_i)$ ~ denoise/estimate.

noisy observation	pixel values
Spectral measurement	MRI image
noisy received bits	bits sent

* Elimination Algorithm for marginalization (exact but can take $O(|\mathcal{X}|^{n})$ operations)

$n = |V|$; $\mathcal{X}$: alphabet of x_i's

$$p(x) = \frac{1}{Z} f_{12}(x_1,x_2) f_{13}(x_1,x_3) f_{25}(x_2,x_5) f_{345}(x_3,x_4,x_5)$$



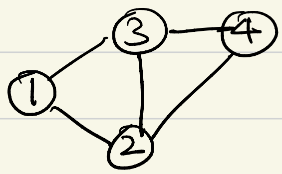
Goal: $p(x_1)$

brute force: $\underbrace{|\mathcal{X}|^4}_{\text{sum}} \cdot \underbrace{|\mathcal{X}|}_{\text{enum. } x_1}$

consider elimination order (5, 4, 3, 2)

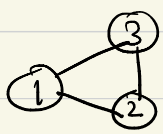
$$p(x_1) = \frac{1}{Z} \sum_{x_2,x_3,x_4} f_{12}(x_1,x_2) f_{13}(x_1,x_3) \underbrace{\sum_{x_5} \boxed{f_{25}(x_2,x_5) f_{345}(x_3,x_4,x_5)}}_{m_5(x_2,x_3,x_4)}$$

$\underbrace{|\mathcal{X}|^3}_{\text{enum.}} \cdot \underbrace{|\mathcal{X}|}_{\text{sum}}$



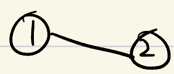
$$= \frac{1}{Z} \sum_{x_2,x_3} f_{12}(x_1,x_2) f_{13}(x_1,x_3) \underbrace{\sum_{x_4} m_5(x_2,x_3,x_4)}_{m_4(x_2,x_3)}$$

$|\mathcal{X}|^2 \cdot |\mathcal{X}|$



$$= \frac{1}{Z} \sum_{x_2} f_{12}(x_1,x_2) \underbrace{\sum_{x_3} f_{13}(x_1,x_3)\, m_4(x_2,x_3)}_{m_3(x_1,x_2)}$$

$|\mathcal{X}|^2 \cdot |\mathcal{X}|$



$$= \frac{1}{Z} \underbrace{\sum_{x_2} f_{12}(x_1,x_2)\, m_3(x_1,x_2)}_{m_2(x_1)}$$

$|\mathcal{X}| \cdot |\mathcal{X}|$

$$m_2 = \underbrace{\begin{bmatrix} \\ \\ \end{bmatrix}}_{|\mathcal{X}|} \longrightarrow p(x_1) = \frac{m_2(x_1)}{\sum_{x_1'} m_2(x_1')}$$

Total: $|\mathcal{X}|^4$ vs. $|\mathcal{X}|^5$

Def. <u>Tree width</u> is $\min_{\text{all orderings}} \{\max$ size of a clique in <u>reconstituted</u> graph$\}$

size of largest vertex set in a tree decomposition.

Def. <u>Reconstituted graph</u> of $G=(V,E)$ w.r.t elimination ordering $\sigma=(\sigma_1 \dots \sigma_n)$

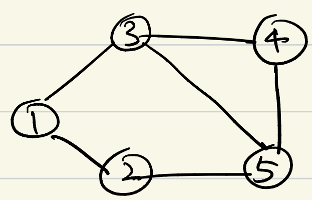
$\tilde{G}=(V,\tilde{E})$

Remove σ_1 and let $E_1 = \{(i,j) \mid (i,j) \in \text{neighbor of } \sigma_1 \text{ in the remaining graph}\}$

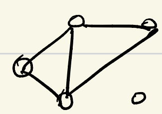
$\vdots$

σ_n

$\tilde{E} = E \cup E_1 \cup E_2 \cdots$

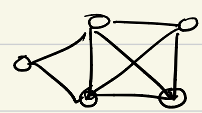


$\sigma = (5, 4, 3, 2, 1)$



E_1

...



$\tilde{E}$

Treewidth $=4$.

---

*Elimination Algorithm

for MAP inference $\arg\max_x P(x) = x^*$

Q. Can we use $\arg\max_{x_1} P_1(x_1) \overset{?}{=} x_1^*$

$$\underset{x_1}{\arg\max} \sum_{x_2 \cdots x_n} P(x_1, x_2^n) \neq \underset{x_1}{\arg\max} \max_{x_2 \cdots x_n} P(x_1, x_2^n)$$

eliminate by maximization with ordering $(5,4,3,2,1)$

$$\underset{x_1 \cdots x_5}{\arg\max} P(x) = \frac{1}{Z} \underset{x_1 \cdots x_4}{\arg\max} f_{12}(x_1, x_2) f_{13}(x_1, x_3) \underbrace{\max_{x_5} f_{25}(x_2, x_5) f_{345}(x_3, x_4, x_5)}_{m_5^*(x_2, x_3, x_4)}$$

$x_5^*(x_2, x_3, x_4) \leftarrow$ maximizer

$\rightarrow x_5^*$

$$= \frac{1}{Z} \arg\max_{x_1^3} f_{12}(x_1,x_2) f_{13}(x_1,x_3) \underbrace{\max_{x_4} m_5^*(x_2,x_3,x_4)}_{m_4^*(x_2,x_3)}$$

$x_4^*(x_2,x_3) \to x_4^*$

$$= \frac{1}{Z} \arg\max_{x_1,x_2} f_{12}(x_1,x_2) \underbrace{\max_{x_3} f_{13}(x_1,x_3)\, m_4^*(x_2,x_3)}_{m_3^*(x_1,x_2)}$$

$x_3^*(x_1,x_2) \to x_3^*$

$$= \frac{1}{Z} \arg\max_{x_1} \underbrace{\max_{x_2} f_{12}(x_1,x_2)\, m_3^*(x_1,x_2)}_{m_2^*(x_1)}$$

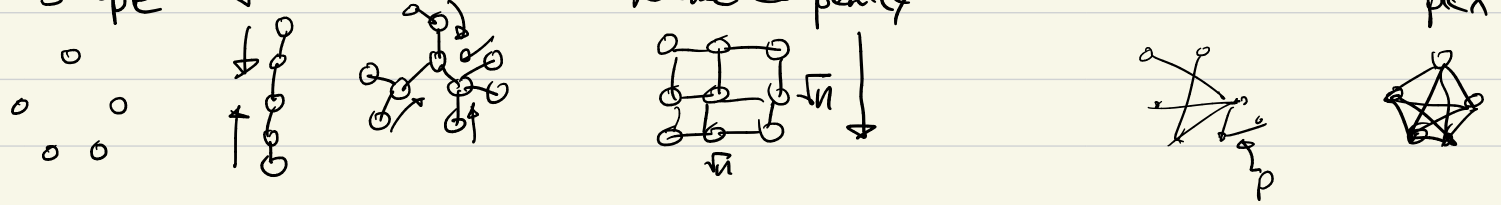
$x_2^*(x_1) \to x_2^* = x_2^*(x_1^*)$

$$= \frac{1}{Z} \arg\max_{x_1} m^*(x_1) \to x_1^*$$

Our first formal notion of "Simplicity" of a graph

Small ⟵ Treewidth $G = ([n], E)$ ⟶ Large

simple ⟵ Inference Complexity $P(x)$ ⟶ Complex



$E=\emptyset$	chain	Tree	2D-Grid	3D Grid	Erdős-Rényi $p=\frac{c}{n}$	
$TW=1$	2	2	$\sqrt{n}$	$n^{2/3}$	n	$TW=n$
$O(n\lvert\mathcal{X}\rvert)$	$\lvert\mathcal{X}\rvert^2$	$\lvert\mathcal{X}\rvert^2$	$\lvert\mathcal{X}\rvert^{\sqrt{n}}$	$\lvert\mathcal{X}\rvert^{n^{2/3}}$	$\lvert\mathcal{X}\rvert^n$	$O(\lvert\mathcal{X}\rvert^n)$
max-degree						
0	2	2	4	6	c	n

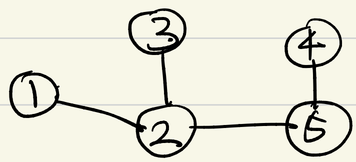
Q. Is E.R graph as hard for inference as K_n?

* Elimination algorithm on a Tree $\Longrightarrow$ Sum-Product algorithm
(loopy) belief propagation

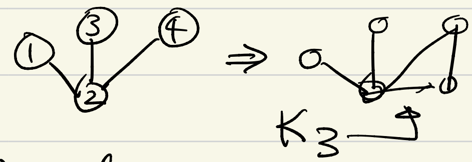
Exact on Tree — Approximate on G

Def. Tree: is a graph
- Connected: $\exists$ path from any i to any j
- no cycles
- $|E| = n-1$

$\Rightarrow$ redundant.



wrong elimination ordering: $(5, 4, 3, 2, 1)$.

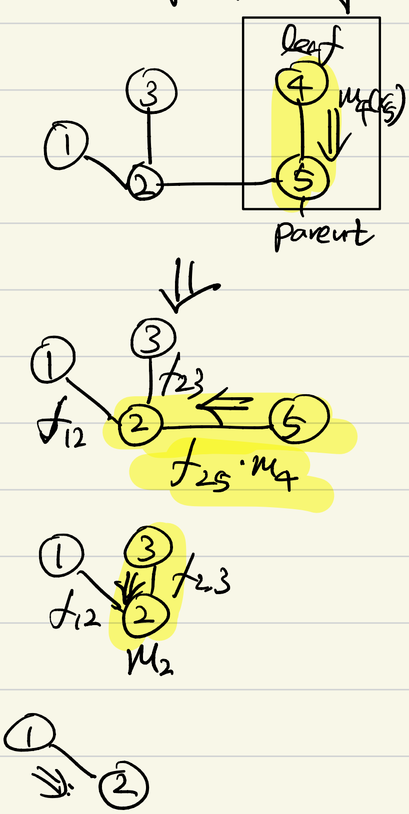


Obvious & Natural elimination ordering: recursively eliminate leaves
$\triangleq$ node with degree 1.

$(4, 5, 3, 1, 2)$.

$$P(x) = \frac{1}{Z} f_{12}(x_1, x_2)\, f_{23}(x_2, x_3)\, f_{25}(x_2, x_5)\, f_{45}(x_4, x_5)$$

message-passing.



$$P(x_2) = \frac{1}{Z} \sum_{x_1 x_3 x_5} f_{12} \cdot f_{23} \cdot f_{25} \cdot \underbrace{\sum_{x_4} f_{45}(x_4, x_5)}_{m_{4\to5}(x_5)}$$: message

$$= \frac{1}{Z} \sum_{x_1 x_3} f_{12} \cdot f_{23} \cdot \underbrace{\sum_{x_5} f_{25}(x_2, x_5) \cdot m_4(x_5)}_{m_{5\to2}(x_2)}$$

$$= \frac{1}{Z} \sum_{x_1} f_{12} \cdot m_5 \underbrace{\sum_{x_3} f_{23}(x_2, x_3)}_{m_{3\to2}(x_2)}$$

$$= \frac{1}{Z} m_{3\to2}(x_2)\, m_{5\to2}(x_2) \underbrace{\sum_{x_1} f_{12}(x_1, x_2)}_{m_{1\to2}(x_2)}$$

$$P(X_2) = \frac{1}{Z} \; m_{3\to 2}(X_2)\, m_{5\to 2}(X_2)\, m_{1\to 2}(X_2)$$