

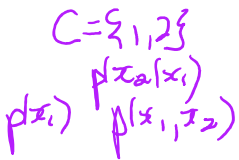
today: • finish UGM
• inference

notes about UGM:

- unlike a DGM, (where could think of C to be (i, π_i) , $\psi_c(x_c) = p(x_i | x_{\pi_i})$)
 $\psi_c(x_c)$ is not directly related to $p(x_c)$
- can rescale any potential without changing joint p i.e. $\psi_c^{new}(x_c) \triangleq \psi_c^{old}(x_c) \cdot A$
(unlike DGM where ψ_c 's were uniquely defined from p)
for some $A > 0$

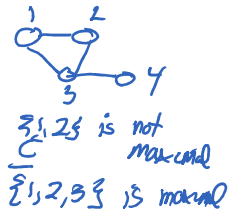
• it is sufficient to consider \mathcal{C}_{max} , the set of maximal cliques

e.g. $C' \in \mathcal{C}$

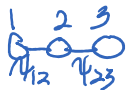
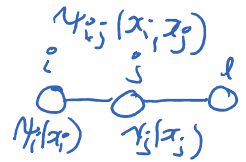


redefine $\psi_c^{new}(x_c) \triangleq \psi_c^{old}(x_c) \cdot \psi_{C'}^{old}(x_{C'})$

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



[we'll see later sometimes convenient to consider "over parameterization" of trees using both $\psi_i(x_i)$ and $\psi_{ij}(x_i, x_j)$



$p(x_2, x_3) \neq \psi_{12}(x_1, x_2) \psi_{23}(x_2, x_3)$
vs. \sum_{x_1}



$p(x_2, x_3) = p(x_2, x_3) p(x_1)$

properties of UGM:

• as before $E \subseteq E' \Rightarrow \mathcal{J}(G) \subseteq \mathcal{J}(G')$

$E = \emptyset \Rightarrow \mathcal{J}(G) =$ set of fully factorized dist.

$E =$ all pairs $\Rightarrow \mathcal{J}(G) =$ all distributions on \mathcal{X}_V
(i.e. G is just one big clique)

• if $\psi_c(x_c) > 0 \forall x_c, \psi_c$ $\langle \theta_c, T_c(x_c) \rangle$

then can write $p(x_v) = \exp(\underbrace{\sum_{c \in \mathcal{C}} \log \psi_c(x_c)} - \log z)$

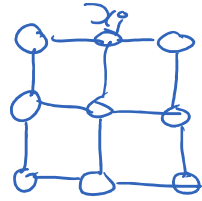
physics link: *negative energy fct.*

e.g. Ising model in physics

$$x_i \in \{0, 1\}$$

node potential $\rightarrow E_i$

edge potential $\rightarrow E_{i,j}$



$$E_i = \psi_i(x_i=1)$$

$$E_{i,j} = \psi_{i,j}(x_i=1, x_j=1)$$

other: social network modeling

$$p(x_A) = \frac{1}{Z} \prod_{c \in \mathcal{C}} \psi_c(x_c)$$

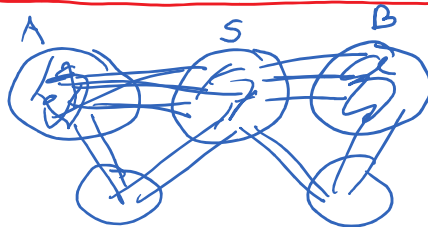
$$p(x_B | x_A) = \prod_{i \in B} p(x_i | x_{N_i})$$

conditional independence for UGM:

def: we say that p satisfies global Markov property (with respect to an undirected graph G)

iff $\forall A, B, S \subseteq V$ s.t. S separates A from B in G
disjoint

then $X_A \perp\!\!\!\perp X_B \mid X_S$



prop:

$p \in \mathcal{G}(G) \Rightarrow p$ satisfies the global Markov property for G

15h33

proof: WLOG, we assume $A \cup B \cup S = V$ } A separated from B by S

$\Gamma \cup \Delta$

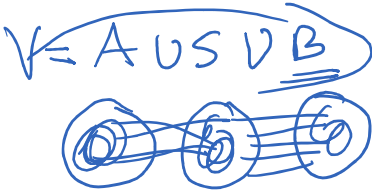
$\vee \parallel \dots$

\rightarrow

proof: WLOG, we assume $A \cup B \cup S = V$ $\} A$ separated from B by S

[why? $X_A \perp\!\!\!\perp X_B \mid X_S$]

$\Rightarrow X_{A'} \perp\!\!\!\perp X_{B'} \mid X_S$ $\forall A' \subseteq A$
 $\forall B' \subseteq B$ by decomposition property
 also A' is separated from B' by S



Let $C \in \mathcal{C}$

cannot have $C \cap A \neq \emptyset$ and $C \cap B \neq \emptyset$
 (because of separation)

thus $p(x) = \frac{1}{2} \prod_{C \in \mathcal{C}} \mathbb{1}_C(x_C) \prod_{C' \in \mathcal{C}} \mathbb{1}_{C'}(x_{C'}) = f(x_{A \cup S}) g(x_{B \cup S})$
 $C \subseteq A \cup S$ $C' \subseteq B \cup S$
 $C' \not\subseteq A \cup S$ and $C' \not\subseteq S$

$p(x_A \mid x_S) \propto p(x_A, x_S) = \sum_{x_B} p(x_V) = \sum_{x_B} f(x_{A \cup S}) g(x_{B \cup S})$
 $= f(x_{A \cup S}) \sum_{x_B} g(x_{B \cup S})$
 (constant w.r. to x_A)

$\Rightarrow p(x_A \mid x_S) = \frac{f(x_{A \cup S})}{\sum_{x_A} f(x_A, x_S)}$

similarly, $p(x_B \mid x_S) = \frac{g(x_{B \cup S})}{\sum_{x_B} g(x_B, x_S)}$

$p(x_A \mid x_S) p(x_B \mid x_S) = \frac{f(x_{A \cup S}) g(x_{B \cup S})}{\sum_{x_A} f(x_A, x_S) \sum_{x_B} g(x_B, x_S)} \} p(x_V) = p(x_A, x_B \mid x_S)$
 $\frac{\sum_{x_A} f(x_A, x_S)}{\sum_{x_A} f(x_A, x_S)} \frac{\sum_{x_B} g(x_B, x_S)}{\sum_{x_B} g(x_B, x_S)} \} \frac{p(x_V)}{p(x_S)}$

i.e. $X_A \perp\!\!\!\perp X_B \mid X_S$ //

$f(x_{A \cup S}) = f(x_A, x_S)$

thm: (Hammersley-Coffard) if $p(x_V) > 0 \forall x_V$

then $\forall C \in \mathcal{C} \setminus \{S\} \rightarrow \exists$ a subset $A \subseteq C \cap A$ and $B \subseteq C \cap B$ such that A and B are separated by S

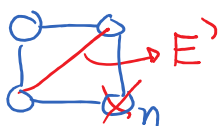
thm: (Hammersley-Clifford) $\text{if } p(x_v) > 0 \forall x_v$

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

proof: see ch 16 of Mike's book; use "Möbius inversion formula" (as exclusion-inclusion principle in proba)

more properties of UGM:

* 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)

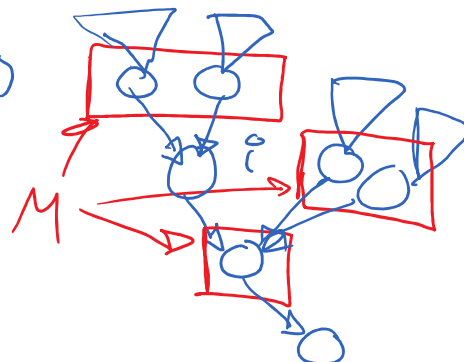
$\sum \text{marginal on } x_{1:m} \text{ for some } p \in \mathcal{J}(G) \} = \mathcal{J}(G')$

DGM vs. UGM

def: Markov blanket for i (for graph G) is the smallest set of nodes M


s.t. $X_i \perp\!\!\!\perp X_V \mid X_M$
"rest"

- for UGM: $M = \{j; \{i,j\} \in E\}$ = set of neighbors of i
- for DGM: $M = \pi_i \cup \underbrace{\text{children}(i)}_{\text{like UGM}} \cup \bigcup_{j \in \text{children}(i)} \pi_j$



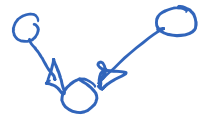
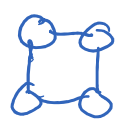
recap:

	DGM	UGM
factorization:	$p(x) = \prod_i p(x_i \pi_i)$	$\frac{1}{Z} \prod_C \psi_C(x_C)$
cond. indep.:	d-separation	separation
marginalization:	not closed in general	closed (connect all neighbors)

marginalization: not closed in general
 e.g. 
 but is fine for leaves

closed
 (connect all neighbors
 of removed node)

cannot
 exactly
 capture some
 families



Moralization:

let G be a DAG; when can we get equivalent UGM?

def: for G a DAG, we call \bar{G} the moralized graph of G

where \bar{G} is an undirected graph with same V

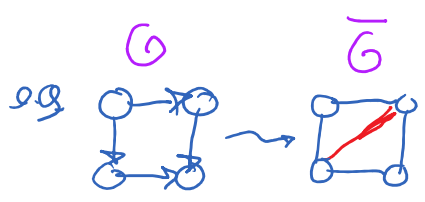
and $\bar{E} = \{ \{i, j\} : (i, j) \in E \}$ {undirected version of E

$\cup \{ \{k, l\} : k \neq l \in \pi_i \text{ for some } i \}$ } "moralization"

connect all the parents of i
 with i in big clique

only needed if $|\pi_i| > 1$
 i.e. v-structure

"marrying the
 parents" (??)



prop: for a dag G with no v-structure [ferst]
 then $\mathcal{I}(G) = \mathcal{I}(\bar{G})$
 DGM UGM

(proof see assignment!)

but in general, can only say $\mathcal{I}(G) \subseteq \mathcal{I}(\bar{G})$

[note that \bar{G} is the minimal undirected graph G'
 s.t. $\mathcal{I}(G) \subseteq \mathcal{I}(G')$]