WASP GM23: Lecture 1

Representation with Undirected Graphs

- In the second introductory lecture we defined two families of distributions associated to a DAG:
  - $\mathbb{X} = [X_1, \ldots, X_m]^T$ a distribution
  - $G = ([m], E)$ a DAG.
  - $\mathcal{M}_F(G) = \{ \mathbb{X} : f_{\mathbb{X}}(x_1, \ldots, x_n) = \prod_{i=1}^{m} f_{X_i | X_{pa_G(i)}}(x_i \mid x_{pa_G(i)}) \}$
  - $\mathcal{M}_{Glo}(G) = \{ \mathbb{X} : \mathbb{X}_A \perp\!\!\!\perp \mathbb{X}_B \mid \mathbb{X}_C$ whenever A, B d-separated given C in $G \}$.
- Our first goal is to show that these two families are equal. To do so, we will first prove analogous results for undirected graphs.
- $G = ([m], E)$ an undirected graph.

$\boxed{\text{Definition}}$ (P) $\mathbb{X}$ satisfies the pairwise Markov property w.r.t. G if

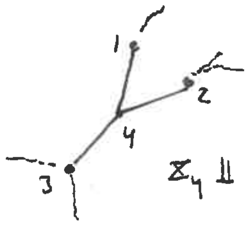
$$X_u \perp\!\!\!\perp X_v \mid \mathbb{X}_{[m] \setminus \{u,v\}} \quad \text{for all } \{u,v\} \notin E.$$

(simply $u \perp\!\!\!\perp v \mid [m] \setminus \{u,v\}$)

$$\mathcal{M}_P(G) = \{ \mathbb{X} : \mathbb{X} \text{ satisfies pairwise MP w.r.t. } G \}.$$

(L) $\mathbb{X}$ satisfies the local Markov Property w.r.t. G if for all $u \in [m]$

$$u \perp\!\!\!\perp [m] \setminus Ne_G[u] \mid Ne_G(u).$$



$X_4 \perp\!\!\!\perp \mathbb{X}_{[m]\setminus\{1,2,3\}} \mid \mathbb{X}$

$$\mathcal{M}_L(G) = \{ \mathbb{X} : \mathbb{X} \text{ satisfies Local MP w.r.t. } G \}.$$

(G) $\mathbb{X}$ satisfies the global Markov property w.r.t. G if for all triples A, B, C of disjoint subsets of $[m]$ such that C separates A and B in G we have $A \perp\!\!\!\perp B \mid C$.

$$\mathcal{M}_{Glo}(G) = \{ \mathbb{X} : \mathbb{X} \text{ satisfies the global MP w.r.t. } G \}.$$

- Our first observation is that these models form a chain of inclusions.

Proposition 1 For an undirected graph (UG) $G = ([m], E)$ and distribution $\mathbb{X} = [X_1, \dots, X_m]^T$ we have $(G) \Rightarrow (L) \Rightarrow (P)$. That is,

$$\mathcal{M}_{Glo}(G) \subseteq \mathcal{M}_L(G) \subseteq \mathcal{M}_P(G).$$

Proof: $(G) \Rightarrow (L)$: Since $Ne_G(u)$ separates u from $[m] \setminus Ne_G[u]$.

$(L) \Rightarrow (P)$: Assume $\mathbb{X} \in \mathcal{M}_L(G)$ and consider $\{u,v\} \notin E$.

It follows that $v \in [m] \setminus Ne_G[u]$.

Want to show: $u \perp\!\!\!\perp v \mid [m] \setminus \{u,v\}$.

Since $\mathbb{X} \in \mathcal{M}_L(G)$ we have $u \perp\!\!\!\perp [m]\setminus Ne_G[u] \mid Ne_G(u)$, else

$$u \perp\!\!\!\perp ([m]\setminus Ne_G[u])\setminus\{v\} \cup \{v\} \mid Ne_G(u)$$

Weak union $\Rightarrow$ $A \perp\!\!\!\perp B \cup D \mid C$ then $A \perp\!\!\!\perp B \mid C \cup D$

So $u \perp\!\!\!\perp v \mid Ne_G(u) \cup ([m]\setminus Ne_G[u])\setminus\{v\}$, or equivalently,

$$u \perp\!\!\!\perp v \mid [m]\setminus\{u,v\}.$$

$\Rightarrow \mathbb{X} \in \mathcal{M}_P(G)$. ■

- If $\mathbb{X}$ satisfies the intersection axiom then we can turn this set containment into an equality...

Definition 2 We say $\mathbb{X}$ satisfies the intersection axiom if whenever $A \perp\!\!\!\perp B \mid C \cup D$ and $A \perp\!\!\!\perp D \mid C \cup B$ then $A \perp\!\!\!\perp B \cup D \mid C$.

Lemma 2 Suppose $f_{\mathbb{X}}(x_1, \dots, x_m) > 0$ for all $[x_1, \dots, x_m]^T$. Then $\mathbb{X}$ satisfies the intersection axiom.

Proof: Exercise.

Example A multivariate normal distribution $\mathbb{X} \sim N(\underline{\mu}, \Sigma)$ satisfies the intersection axiom.

- The following theorem holds for all positive distributions, including multivariate normal distributions.

$\boxed{\text{Theorem } ④}$ (Pearl and Paz; 1987) If $\mathbb{X} = [X_1, \dots, X_m]^T$ satisfies the intersection axiom then $\mathbb{X} \in M_{Glo}(G)$ if and only if $\mathbb{X} \in M_L(G)$ if and only if $\mathbb{X} \in M_P(G)$. ~~But~~

Proof: By Proposition ① it suffices to show $M_P(G) \subseteq M_{Glo}(G)$.

- S'pose $\mathbb{X} \in M_P(G)$ and C separates A and B in G with $A, B \neq \emptyset$.
- We show $A \perp\!\!\!\perp B \mid C$ by backwards induction on $n = |C|$.

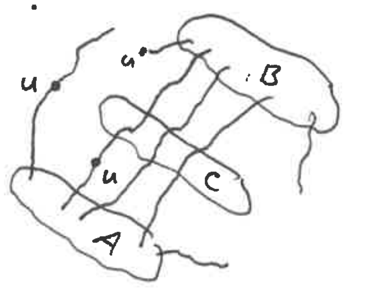
Base: If $n = m-2$ then A, B singletons so $A \perp\!\!\!\perp B \mid C$ is exactly a statement $u \perp\!\!\!\perp v \mid [m] \setminus \{u, v\}$. exactly the statements implied by $\mathbb{X} \in M_P(G)$. ✓

Inductive Hypothesis: Assume result holds $\forall A, B, C$ with $|C| > n$ and consider a separation statement for C with $|C| = n < m-2$.

① If $[m] = A \cup B \cup C$ then without loss of generality $|A| > 1$. For $u \in A$ we can apply the weak union property for graph separation as follows:

$$A \perp_G B \mid C \iff A \setminus \{u\} \cup \{u\} \perp_G B \mid C$$
$$\Downarrow$$
$$\underbrace{A \setminus \{u\} \perp_G B \mid C \cup \{u\} \quad \text{and} \quad u \perp_G B \mid C \cup A \setminus \{u\}}$$
$$\Downarrow \text{ IH since } |C \cup \{u\}|, |C \cup A \setminus \{u\}| > n$$
$$\underbrace{A \setminus \{u\} \perp\!\!\!\perp B \mid C \cup \{u\} \quad \text{and} \quad u \perp\!\!\!\perp B \mid C \cup A \setminus \{u\}}$$
$$\Downarrow \text{(Intersection axiom)}$$
$$A \perp\!\!\!\perp B \mid C. ✓$$

② If $[m] \supsetneq A \cup B \cup C$ choose $u \in [m] \setminus A \cup B \cup C$.
Then $C \cup \{u\}$ separates A and B in G
(I.H.) $\Rightarrow A \perp\!\!\!\perp B \mid C \cup \{u\}$.



- Moreover either $A \cup C$ separates B from u or $B \cup C$ separates A from u. Otherwise, we have a path from A to B through u avoiding C.

  $\rightsquigarrow$ Without loss of generality assume first case.

$$\text{(I. H.)} \Rightarrow u \perp\!\!\!\perp B \mid C \cup A$$

$$\text{(Intersection)}: u \perp\!\!\!\perp B \mid C \cup A \text{ and } A \perp\!\!\!\perp B \mid C \cup \{u\}$$
$$\Rightarrow A \cup \{u\} \perp\!\!\!\perp B \mid C$$

$$\text{(Decomposition)}: A \cup \{u\} \perp\!\!\!\perp B \mid C \Rightarrow A \perp\!\!\!\perp B \mid C. \quad \blacksquare$$

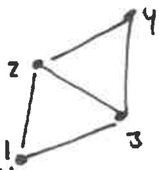
- We now know that for all positive distributions these three families of distributions defined by conditional independence relations are all equal. It is also useful to understand how ~~distribution~~ the pdf of distributions in these models factor.

Factorization Criteria for Undirected Graphical Models

Definition ③ $\mathbb{X} = [\mathbb{X}_1, \dots, \mathbb{X}_m]^T$ factorizes w.r.t. a UG $G = ([m], E)$ if for all cliques $A \subseteq [m]$ in G there exist nonnegative functions $\psi_A(x_A)$ such that

$$f_{\mathbb{X}}(x_1, \dots, x_m) = \prod_{\substack{A : A \text{ a clique} \\ \text{in } G}} \psi_A(x_A).$$

Note The functions $\psi_A(x_A)$ are not uniquely determined.



$$f_{\mathbb{X}}(x) = \psi_{123}(x_{1,2,3})\,\psi_{1,2}(x_{1,2})\,\psi_{1,3}(x_{1,3})\,\underbrace{\psi_{2,3}(x_{2,3})}_{=\left(\sqrt{\psi_{2,3}(x_{2,3})}\right)\left(\sqrt{\psi_{2,3}(x_{2,3})}\right)}\,\psi_{3,4}(x_{3,4})\,\psi_{2,4}(x_{2,4})\,\psi_{2,3,4}(x_{2,3,4})$$

$$\underbrace{\psi_{123}(x_{1,2,3})\,\psi_{1,2}(x_{1,2})\,\psi_{1,3}(x_{1,3})\left(\sqrt{\psi_{2,3}(x_{2,3})}\right)}_{=\tilde\psi_{1,2,3}(x_{1,2,3})}\underbrace{\left(\sqrt{\psi_{2,3}(x_{2,3})}\right)\psi_{3,4}(x_{3,4})\,\psi_{2,4}(x_{2,4})\,\psi_{2,3,4}(x_{2,3,4})}_{=\tilde\psi_{2,3,4}(x_{2,3,4})}$$

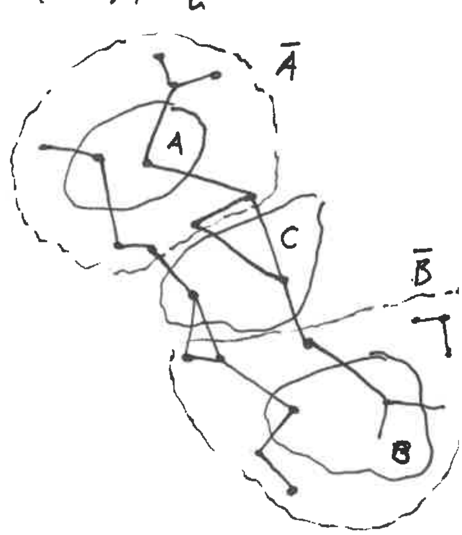
$$= \tilde\psi_{1,2,3}(x_{1,2,3})\,\tilde\psi_{2,3,4}(x_{2,3,4}) \prod_{\{i,j\} \in E} \tilde\psi_{i,j}(x_{i,j}) \quad \text{where } \tilde\psi_{i,j}(x_{i,j}) = 1$$

$\Rightarrow$ Without loss of generality, we can work with only maximal cliques in G : $\mathcal{C}(G)$

$$f_{\mathbb{X}}(x_1,\dots,x_m) = \prod_{C\in\mathcal{C}(G)} \psi_C(x_C).$$

Proposition ② If $\mathbb{X}=[X_1,\dots,X_m]^T$ factors w.r.t. the UG $G=([m],E)$ then $\mathbb{X}\in\mathcal{M}_{Glo}(G)$.

Proof: S'pose $\mathbb{X}$ factors w.r.t. G and that $A\perp_G B\mid C$.



- $\bar{A}$ = connected components in $G\setminus C$ containing A
- $\bar{B} = [m]\setminus(\bar{A}\cup C)$
- Since C separates A and B their element are in different connected components of $G\setminus C$

$\Rightarrow$ any maximal clique in G is in either $\bar{A}\cup C$ or $\bar{B}\cup C$.

- $\mathcal{C}_A$ = maximal cliques in $\bar{A}\cup C$.

$$\Rightarrow f_{\mathbb{X}}(x) = \prod_{D\in\mathcal{C}(G)} \psi_D(x_D) = \left(\prod_{D\in\mathcal{C}_A}\psi_D(x_D)\right)\left(\prod_{D\in\mathcal{C}(G)\setminus\mathcal{C}_A}\psi_D(x_D)\right) = h(x_{\bar{A}\cup C})\,k(x_{\bar{B}\cup C}).$$

$\Rightarrow \bar{A}\perp\!\!\!\perp\bar{B}\mid C. \Longrightarrow A\perp\!\!\!\perp B\mid C$ by Decomposition $\boxtimes$

Exercise: $\mathbb{X}_A\perp\!\!\!\perp\mathbb{X}_B\mid\mathbb{X}_C$ if and only if $f(x_A,x_B,x_C)=h(x_A,x_C)k(x_B,x_C)$ for some functions h and k.

- $\mathcal{M}_F(G)=\{\mathbb{X} : \mathbb{X}$ factors w.r.t. $G\}$. We have

$$\mathcal{M}_F(G)\subseteq\mathcal{M}_{Glo}(G)\subseteq\mathcal{M}_L(G)\subseteq\mathcal{M}_P(G).$$

- The Hammersley-Clifford theorem says these sets are all equal if we restrict to positive, continuous distributions:

$\boxed{\text{Theorem 5}}$ (Hammersley - Clifford Theorem) Let $\mathbb{X}$ have a positive and continuous density. Then $\mathbb{X} \in M_F(G)$ iff and only if $\mathbb{X} \in M_P(G)$. That is,

$$(F) \iff (G) \iff (L) \iff (P).$$

Proof: Lauritzen Theorem 3.9.

- So far we have obtained equalities amongst these properties by placing restrictions on the distributions. Alternatively, we can obtain equalities by restricting the choice of graph G.

$\boxed{\text{Proposition 3}}$ Let $G = ([m], E)$ be a chordal graph. $\mathbb{X} = [X_1, \dots, X_m]^T$ factorizes according to G if and only if $\mathbb{X}$ satisfies the global M.P. w.r.t. G. That is, $(F) \iff (G)$.

Proof: Since G is chordal it has a proper weak decomposition.

We have the following fact:

$\boxed{\text{Fact}}$ (~~Proposition~~ Proposition 3.17, Lauritzen) If (A, B, C) a proper weak decomposition of G then $\mathbb{X} \in M_{Glo}(G) \iff \mathbb{X}_{A\cup C}$ and $\mathbb{X}_{B\cup C}$ satisfy the global M.P. w.r.t. the induced subgraphs $G_{A\cup C}$ and $G_{B\cup C}$, respectively.

Moreover

$$f_{\mathbb{X}}(x) f_{\mathbb{X}_C}(x_C) = f_{\mathbb{X}_{A\cup C}}(x_{A\cup C}) f_{\mathbb{X}_{B\cup C}}(x_{B\cup C}).$$

Using the fact, we induct on $|\mathcal{C}(G)|$.

Base: $|\mathcal{C}(G)| = 1$ trivial since G complete.

I.H.: Assume true for all UGs with at most n maximal cliques.

If G has $|\mathcal{C}(G)| = n+1$ and $\mathbb{X} \in M_{Glo}(G)$, then since G has the proper weak decomposition (A, B, C) then $\mathbb{X}_{A\cup C} \in M_{Glo}(G_{A\cup C})$ and $\mathbb{X}_{B\cup C} \in M_{Glo}(G_{B\cup C})$.

(I. H.) $\Rightarrow$ $f_{\mathbb{X}_{A\cup C}}(x_{A\cup C})$ and $f_{\mathbb{X}_{B\cup C}}(x_{B\cup C})$ factorize according to $G_{A\cup C}$ and $G_{B\cup C}$, respectively.

$\Rightarrow$ $f_{\mathbb{X}_{A\cup C}}(x_{A\cup C})$ contains a factor $\psi_C(x_C)$.

$$\Rightarrow f_{\mathbb{X}}(x) = \frac{f_{\mathbb{X}_{A\cup C}}(x_{A\cup C})\, f_{\mathbb{X}_{B\cup C}}(x_{B\cup C})}{f_{\mathbb{X}_C}(x_C)}$$ factorizes w.r.t. G. ∎

- Note that the distributions (discrete or continuous) that we consider are positive on their spaces of possible outcomes (otherwise the outcomes aren't really possible). So combining $\boxed{\text{Prop } 3}$ with $\boxed{\text{Theorem } 4}$ means that for G chordal we have

$$(F) \Leftrightarrow (G) \Leftrightarrow (L) \Leftrightarrow (P).$$