ECE 175B: Probabilistic Reasoning and Graphical Models: Markov Properties and Hammersley-Clifford Theorem

Florian Meyer

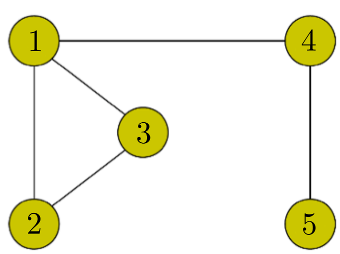
Electrical and Computer Engineering Department
University of California San Diego

UC San Diego
JACOBS SCHOOL OF ENGINEERING

0

Undirected Graphs

- Let $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ be an undirected graph with nodes $\mathcal{V}$ and edges $\mathcal{E}$
- *Example:*



$$\mathcal{V} = \{1, 2, 3, 4, 5\}$$

$$\mathcal{E} = \{(1,2), (1,3), (1,4), (4,5)\}$$

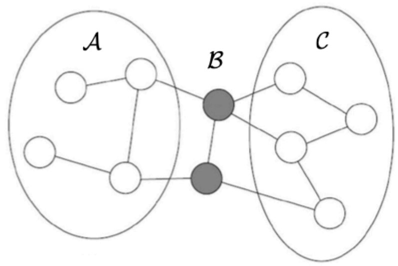
- Let $\mathcal{X}_\mathcal{G} = \{\boldsymbol{x}_v\}_{v \in \mathcal{V}}$ be a set of random variables

1

1

Global Markov Property

- Let $\mathcal{G}$ be an undirected graph with subgraphs $\mathcal{A}, \mathcal{B}, \mathcal{C}$



- $\mathcal{B}$ separates $\mathcal{A}$ and $\mathcal{C}$ if every path from a node in $\mathcal{A}$ to a node in $\mathcal{C}$ passes through a node in $\mathcal{B}$:

$$< \mathcal{A}|\mathcal{B}|\mathcal{C} >_{\mathrm{d}}$$

- The global Markov property is satisfied if for any disjoint $\mathcal{A}, \mathcal{B}, \mathcal{C}$ such that $\mathcal{B}$ separates $\mathcal{A}$ and $\mathcal{C}$, $\boldsymbol{x}_{\mathcal{A}}$ is independent of $\boldsymbol{x}_{\mathcal{C}}$ given $\boldsymbol{x}_{\mathcal{B}}$: $< \mathcal{A}|\mathcal{B}|\mathcal{C} >_{\mathrm{d}} \implies \boldsymbol{x}_{\mathcal{A}} \perp\!\!\!\perp \boldsymbol{x}_{\mathcal{C}} \,|\, \boldsymbol{x}_{\mathcal{B}}$

2

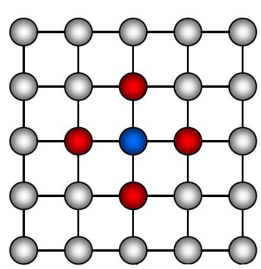
2

Local Markov Property

- For each node $\boldsymbol{x}_v \in \mathcal{X}$, there is a unique Markov blanket denoted $\mathrm{mb}_{\boldsymbol{x}_v}$, which is the set of random variable corresponding to neighbors of v in $\mathcal{G}$ (nodes that share an edge with v)

- Definition: The local Markov independencies associated with $\mathcal{G}$ are given by

$$\boldsymbol{x}_v \perp\!\!\!\perp \mathcal{X}_{\mathcal{G}} \setminus \{\boldsymbol{x}_v\} \setminus \mathrm{mb}_{\boldsymbol{x}_v} | \mathrm{mb}_{\boldsymbol{x}_v}, \quad \forall \boldsymbol{x}_v \in \mathcal{X}_{\mathcal{G}}$$



3

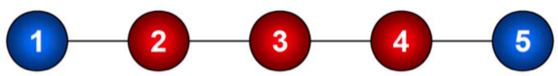
3

Pairwise Markov Property

- Any two variables $\boldsymbol{x}_v, \boldsymbol{x}_u$ corresponding to non-adjacent nodes v, u in graph $\mathcal{G}$ are conditionally independent given all the other variables, i.e.,

$$\boldsymbol{x}_v \perp\!\!\!\perp \boldsymbol{x}_u \,\big|\, \mathcal{X}_{\mathcal{G}} \backslash \{\boldsymbol{x}_v, \boldsymbol{x}_u\}, \quad (u, v) \notin \mathcal{E}$$

- Example: $\boldsymbol{x}_1 \perp\!\!\!\perp \boldsymbol{x}_5 \,\big|\, \{\boldsymbol{x}_2, \boldsymbol{x}_3, \boldsymbol{x}_4\}$



$$\mathcal{V} = \{1, 2, 3, 4, 5\}$$

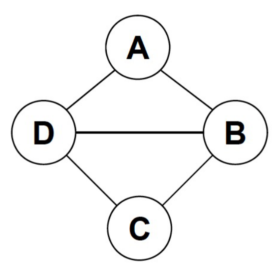
$$\mathcal{E} = \{(1,2), (2,3), (3,4), (4,5)\}$$

4

4

Cliques

- For $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, a complete subgraph or clique is a subgraph $\mathcal{G}_1 = (\mathcal{V}_1 \subseteq \mathcal{V}, \mathcal{E}_1 \subseteq \mathcal{E})$ with fully interconnected nodes $\mathcal{V}_1$
- A maximal clique is complete subgraph $\mathcal{G}_1$ of $\mathcal{G}$ where any graph $\mathcal{G}_2 = (\mathcal{V}_1 \subset \mathcal{V}_2 \subseteq \mathcal{V}, \mathcal{E}_1 \subset \mathcal{E}_2 \subseteq \mathcal{E})$ is not complete
- A sub-clique is a clique that is not maximal
- Example:
  - maximal cliques: $\{A, B, D\}$ and $\{B, C, D\}$
  - sub-cliques: $\{A, B\}$, $\{B, C\}$, …



5

5

Factorization

- The distribution $p(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_{|\mathcal{V}|})$ factorizes according to graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ if its density can be written in the form

$$p(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_{|\mathcal{V}|}) = \frac{1}{Z} \prod_{c \in \mathcal{C}} \psi_c(\boldsymbol{x}^{(c)})$$

where the potentials $\psi_c(\boldsymbol{x}^{(c)})$ are non-negative functions associated with cliques $\mathcal{C}$ of $\mathcal{G}$ and Z is the partition function

$$Z = \sum_{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_{|\mathcal{V}|}} \prod_{c \in \mathcal{C}} \psi_c(\boldsymbol{x}^{(c)})$$

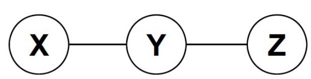
- The potential functions $\psi_c(\boldsymbol{x}^{(c)})$ can be understood as contingency functions of its arguments or local building blocks -> often no probabilistic interpretation

- Factorization according to $\mathcal{G}$ implies the global Markov property

6

6

Interpretation of Clique Potentials



- This model implies that $\boldsymbol{x} \perp\!\!\!\perp \boldsymbol{z} \,|\, \boldsymbol{y}$
- Thus, the joint distribution must factorize as

$$p(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}) = p(\boldsymbol{y}) p(\boldsymbol{x}|\boldsymbol{y}) p(\boldsymbol{z}|\boldsymbol{y})$$

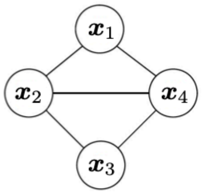
- We can also write this as $p(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}) = p(\boldsymbol{y}, \boldsymbol{x}) p(\boldsymbol{z}|\boldsymbol{y})$ or $p(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}) = p(\boldsymbol{z}, \boldsymbol{y}) p(\boldsymbol{x}|\boldsymbol{y})$
  – cannot have all potentials to be marginals
  – cannot have all potentials to be conditionals

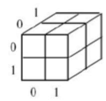
- Non-negative clique potentials can be thought of as general ``compatibility'', ``goodness'' or ``happiness'' functions over their variables, but not as probability distributions

7

7

Example: Factorization Using Maximal Cliques





$\psi_c(\cdot,\cdot,\cdot)$

$$p(\boldsymbol{x}_1,\boldsymbol{x}_2,\boldsymbol{x}_3,\boldsymbol{x}_4) = \frac{1}{Z}\,\psi_{124}(\boldsymbol{x}_1,\boldsymbol{x}_2,\boldsymbol{x}_4)\,\psi_{234}(\boldsymbol{x}_2,\boldsymbol{x}_3,\boldsymbol{x}_4)$$

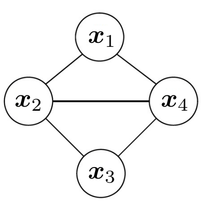
$$Z = \sum_{\boldsymbol{x}_1,\boldsymbol{x}_2,\boldsymbol{x}_3,\boldsymbol{x}_4} \psi_{124}(\boldsymbol{x}_1,\boldsymbol{x}_2,\boldsymbol{x}_4)\,\psi_{234}(\boldsymbol{x}_2,\boldsymbol{x}_3,\boldsymbol{x}_4)$$

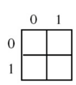
- $p(\boldsymbol{x}_1,\boldsymbol{x}_2,\boldsymbol{x}_3,\boldsymbol{x}_4)$ can be represented as two 3-D tables instead of one 4-D table
- The factorization using maximal cliques
  - can always be used without loss of generality
  - often obscures structure that is present in the original set of potentials

8

8

Example: Factorization Using Sub-Cliques





$\psi_c(\cdot,\cdot)$

$$p'(\boldsymbol{x}_1,\boldsymbol{x}_2,\boldsymbol{x}_3,\boldsymbol{x}_4) = \frac{1}{Z}\,\psi_{12}(\boldsymbol{x}_1,\boldsymbol{x}_2)\,\psi_{14}(\boldsymbol{x}_1,\boldsymbol{x}_4)\psi_{23}(\boldsymbol{x}_2,\boldsymbol{x}_3) \times \psi_{24}(\boldsymbol{x}_2,\boldsymbol{x}_4)\psi_{34}(\boldsymbol{x}_3,\boldsymbol{x}_4)$$

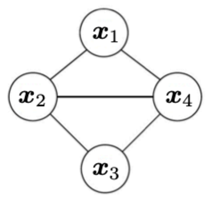
$$Z = \sum_{\boldsymbol{x}_1,\boldsymbol{x}_2,\boldsymbol{x}_3,\boldsymbol{x}_4} \psi_{12}(\boldsymbol{x}_1,\boldsymbol{x}_2)\,\psi_{14}(\boldsymbol{x}_1,\boldsymbol{x}_4)\,\psi_{23}(\boldsymbol{x}_2,\boldsymbol{x}_3) \times \psi_{24}(\boldsymbol{x}_2,\boldsymbol{x}_4)\,\psi_{34}(\boldsymbol{x}_3,\boldsymbol{x}_4)$$

- $p'(\boldsymbol{x}_1,\boldsymbol{x}_2,\boldsymbol{x}_3,\boldsymbol{x}_4)$ can be represented as five 2-D tables instead of one 4-D table
- Markov networks with pairwise interactions is a widely used special case

9

9

Example: Canonical Factorization



$$p''(\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_3, \boldsymbol{x}_4) = \frac{1}{Z}\,\psi_{124}(\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_4)\,\psi_{234}(\boldsymbol{x}_2, \boldsymbol{x}_3, \boldsymbol{x}_4)$$
$$\times\,\psi_{12}(\boldsymbol{x}_1, \boldsymbol{x}_2)\psi_{14}(\boldsymbol{x}_1, \boldsymbol{x}_4)\psi_{23}(\boldsymbol{x}_2, \boldsymbol{x}_3)\psi_{24}(\boldsymbol{x}_2, \boldsymbol{x}_4)\psi_{34}(\boldsymbol{x}_3, \boldsymbol{x}_4)$$
$$\times\,\psi_1(\boldsymbol{x}_1)\psi_2(\boldsymbol{x}_2)\psi_3(\boldsymbol{x}_3)\psi_4(\boldsymbol{x}_4)$$

$$Z = \sum_{\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_3, \boldsymbol{x}_4} \psi_{123}(\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_4)\,\psi_{234}(\boldsymbol{x}_2, \boldsymbol{x}_3, \boldsymbol{x}_4)\ldots$$

- Most general factorization that subsumes any other factorization according to $\mathcal{G}$ as special case

10

Hammersley-Clifford Theorem

- A positive distribution $p(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_{|\mathcal{V}|})$ satisfies the pairwise Markov property with respect to graph $\mathcal{G}$ if and only if it factorizes according to

$$p(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_{|\mathcal{V}|}) = \frac{1}{Z}\prod_{c \in \mathcal{C}} \psi_c(\boldsymbol{x}^{(c)})$$

  where $\mathcal{C}$ are cliques of $\mathcal{G}$ and Z is the partition function

- The Hammersley-Clifford theorem
  - identifies weak assumptions on distributions so that equivalence holds between Markov properties and factorization
  - is the central results of the theory of undirected graphical models

11

Markov Properties and Factorization

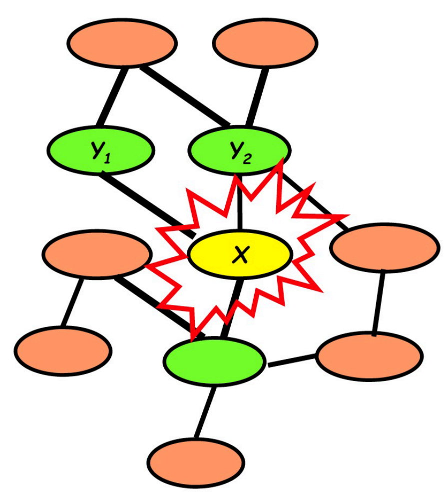
- Factorization $(\boldsymbol{F})$ of a distribution $p(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_{|\mathcal{V}|})$ according to $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ implies global $(\boldsymbol{G})$, local $(\boldsymbol{L})$, and pairwise $(\boldsymbol{P})$ Markov properties, i.e., $(\boldsymbol{F}) \Rightarrow (\boldsymbol{G}) \Rightarrow (\boldsymbol{L}) \Rightarrow (\boldsymbol{P})$
- However, in general $(\boldsymbol{P}) \not\Rightarrow (\boldsymbol{L}) \not\Rightarrow (\boldsymbol{G}) \not\Rightarrow (\boldsymbol{F})$
- Hammersley and Clifford showed that for (strictly) positive density functions $(\boldsymbol{P}) \Rightarrow (\boldsymbol{F})$ and thus $(\boldsymbol{P}) \Leftrightarrow (\boldsymbol{L}) \Leftrightarrow (\boldsymbol{G}) \Leftrightarrow (\boldsymbol{F})$

12

12

Summary: Undirected Graphical Models

- Markov Properties: Conditional independency statements can be extracted from the graph
- Factorization: Local contingency functions (potentials) for each cliques in the graph completely determine the joint distribution
- Hammersley-Clifford theorem: For strictly positive distributions equivalence holds between Markov properties and factorization



13

13