Lecture 4 From Directed to Undirected Graphical Models

Xuanxi Zhang

Topics:

- Gibbs distributions / Markov Models
- Existence and Uniqueness of Markov Models
- Directed vs Undirected Graphical Models

Recall from Lecture 3: Bayesian Nets / Directed Graphical Models

$$P(X_1, X_d) = \prod_i P(X_i \mid X_{A(i)})$$

Pros:

- **Efficient:** Generative process along topological order.
- **Self-normalised:** Local factors are probabilities, so the joint model is automatically normalised.

Cons:

- Conditional independencies are not explicit (rely on d-separation!).
- Lack of existence & uniqueness.

Can we alleviate some of these issues?

1 Undirected Graph Representation

Consider a graph $G = (V, E)$ with $V = \{X_1, \ldots, X_n\}$ (random variables) and E (undirected edges)

- In the directed case, we first build a factorization of the joint probability, i.e.,

$$P(X_1, \ldots, X_n) = \prod_i P(X_i \mid X_{A(i)})$$

  Then, we "draw" an edge $j \to i$ whenever $j \in A(i)$.

- In the undirected case, first, we postulate that:

$$X \perp Y \mid Z$$

  whenever nodes in Z separate (topologically) node X from Y on G. This represents the **Markov property** on G.

Conceptual Relationship

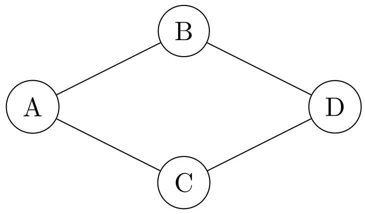
- Undirected Graphical Models starts with a **Conditional Independence** to define the graph, then obain the **Factorization** from the graph.
- Directed Graphical Models starts with a **Factorization** to define the graph, then obtain the **Conditional Independence** from the graph.

Question: Are directed and undirected graphical models expressing the same **class of distributions**? **NO!**

Example 1: Given Conditional Independencies

$$A \perp D \mid \{B, C\}$$

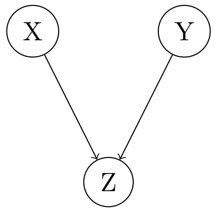
$$B \perp C \mid \{A, D\}$$



Can we build a directed graphical model consistent with these conditional independencies?

Example 2: Given

1. $X \perp Y$
2. $X \not\perp Y \mid Z$



Issue: Constructing a undirected model leads to failures:

- If $X - Z - Y$, then (2) fails!
- If full connected, then (1) fails!

Conclusion: Bayesian Nets (Directed Graphical Models) and Graphical Models (Undirected Graphical Models) have intersection but **not** equivalent!

- BN: Models with "canonical" topological order.
- GM: Models with some *exchangeability*.

2 Parameterization of GM and Gibbs Distributions

Recall BN Factorization:

$$P(X_1, \ldots, X_n) = \prod_i P(X_i \mid X_{A(i)})$$

"Building blocks" are **conditional probabilities**. This form compatible with topological order; however, **undirected models aim to remove this constraint**.

We still want a **local factorization** of the joint density (in order to beat the curse of dimensionality (CoD)). Let start with following formulation:

$$P(X_1, \ldots, X_n) = \prod_{C \in \mathcal{C}} \psi_C(X_C),$$

where $\mathcal{C} \subseteq \mathcal{P}(\{1, 2, \cdots, n\})$ to be determined. $X_C = \{X_i \mid i \in C\}$ and ψ_C is a potential function locally on X_C (usually not a probability distribution).

We want this factorization to be compatible with the **Markov property** on G, which means X_i and X_j are conditionally independent given the rest if i, j are not neighbors in the graph, which means

$$P\left(X_i, X_j \mid \{X_k\}_{k \neq i,j}\right) = F(X_i) G(X_j)$$

Some calculation

$$P\left(X_i, X_j \mid \{X_k\}_{k \neq i,j}\right) = \frac{P(X_1, \ldots, X_n)}{\int P(X_1, \ldots, X_n)\, dX_i\, dX_j} = \frac{\prod_{C \in \mathcal{G}} \psi_C(X_C)}{\int \prod_{C \in \mathcal{G}} \psi_C(X_C)\, dX_i\, dX_j}$$

Expanding the terms,

$$\prod_{C\in\mathcal{G}}\psi_C(X_C)=\prod_{(i,j)\in C}\psi_C(X_C)\cdot\prod_{j\in C,i\notin C}\psi_C(X_C)\cdot\prod_{i\in C,j\notin C}\psi_C(X_C)\cdot\prod_{i\notin C,j\notin C}\psi_C(X_C)$$

Taking the integral leads to

$$P\left(X_i,X_j\mid\{X_k\}_{k\neq i,j}\right)=\frac{\prod_{(i,j)\in C}\psi_C(X_C)}{\int\prod_{(i,j)\in C}\psi_C(X_C)dX_idX_j}\cdot\frac{\prod_{j\in C,i\notin C}\psi_C(X_C)}{\int\prod_{j\in C,i\notin C}\psi_C(X_C)dX_j}\cdot\frac{\prod_{i\in C,j\notin C}\psi_C(X_C)}{\int\prod_{i\in C,j\notin C}\psi_C(X_C)dX_i}$$

We want this function to be of the form: $F(X_i)\cdot G(X_j)$

- $\Rightarrow (i,j)$ **cannot belong to any** C
- $\Rightarrow C$ only contains nodes X_i that are connected with each other.
- $\Rightarrow \mathcal{C}$ only contains **cliques** (fully connected subgraph) of G.
- Since a clique C contains all smaller cliques $C'\subset C$, we can reduce ourselves to the set $\mathcal{G}$ of **maximal cliques**.
- C is a **maximal clique** if $C\cup\{x_i\}$ is not a clique for all $i\notin C$.

2.1 Summary

$$\mathcal{C}=\{C\mid C \text{ is a maximal clique of } G\}$$

$$\psi_C(X_C) \text{ is an arbitrary non-negative potential.}$$

Probability distribution is parameterized as

$$p(x)=\frac{1}{Z}\prod_{C\in\mathcal{C}}\psi_C(X_C),$$

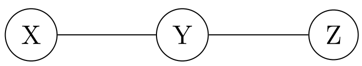
where the partition function is:

$$Z=\int\left(\prod_C\psi_C(X_C)\right)dx.$$

We say that P is a **Gibbs distribution** that factorizes over G.

2.2 Meaning of the local potentials

Example:



Given the independence:

$$X\perp Z\mid Y$$

we can factorize:

$$P(X,Y,Z)=P(X\mid Y)P(Y)P(Z\mid Y)$$

Rewriting,

$$P(X,Y,Z)=P(X\mid Y)P(Y)^{\alpha}P(Y)^{1-\alpha}P(Z\mid Y)$$

which gives us:

$$\psi_1(X,Y),\quad \psi_2(Y,Z)$$

where ψ_1,ψ_2 are **not** probability distributions.
General Case:

$$P(X,Y,Z)\neq P(X,Y)P(Y,Z).$$

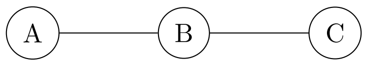
3 From Graph to Markov Property

We start form this question: given a graph G, can we list all conditional independencies invovling a given point X?

Definition 1 (Markov Blanket)**.** A set $A \subseteq \mathcal{X}$ is a **Markov Blanket** for X if $X \notin A$, and A is a minimal set such that:

$$X \perp \mathcal{X} \setminus A \cup \{X\} \mid A.$$

The **Markov Blanket** in **undirected graphical models** is precisely the set of neighbors in G! Consider following example:



with Gibbs Model

$$P(a,b,c) = \frac{1}{Z}\psi_1(a,b) \cdot \psi_2(b,c)$$

Lets verifying the markov Pproperty $A \perp C \mid B$.

$$P(a,c \mid b) = \frac{P(a,b,c)}{P(b)} = \frac{\psi_1(a,b) \cdot \psi_2(b,c)}{Z \cdot P(b)} = \frac{\psi_1(a,b) \cdot \psi_2(b,c)}{\int \psi_1(a,b) \cdot \psi_2(b,c)\, da\, dc} = \frac{\psi_1(a,b)}{\int \psi_1(a,b)\, da} \cdot \frac{\psi_2(b,c)}{\int \psi_2(b,c)\, dc}$$

This is an instance of a more general property:

Theorem 1 (K & F, Theorem 4.1)**.** If P is a **Gibbs distribution** factorizing over G, then G is an $\mathcal{I}$-map for P, i.e.,

$$\mathcal{I}(G) \subseteq \mathcal{I}(P)$$

Proof Sketch If Y separates X and Z, then there are no direct edges between X and Z. Any clique is either in $X \cup Y$ or in $Z \cup Y$. Thus,

$$P(X_1, \ldots, X_n) = \frac{1}{Z}\psi_1(X,Y) \cdot \psi_2(Z,Y)$$

which reduces to the **previous example**. □

4 From Markov Property to Graph

In last section, we showed that given a graph G(and corresponding gibbs distribution), we can deduce the Markov property. Can we deduce that P is Gibbs w.r.t G just from the Markov property?

Theorem 2 (Hammersley-Clifford)**.** Let $P > 0$ over X, and let G be an **undirected graph** over X. If G is an $\mathcal{I}$-map over P, then P is a **Gibbs distribution** w.r.t. G.

Proof: Recitation

Interpretation: The Hammersley-Clifford theorem gives us an **equivalence** between two sources of structure:

- **Factorization** (expressed at the level of the density)

  $$P(X_1, \ldots, X_n) = \frac{1}{Z}\prod_C \psi_C(X_C)$$

  (Gibbs Distribution)

- **Independence** (expressed at the level of random variables)

  $$X_i \perp X_j \mid X_k, k \notin \{i,j\}$$

  (Markov Assumption)

Remark: Positivity assumption is necessary! (We'll see a counter-example in HW 2.)