

From Distributions to Markov Networks

Sargur Srihari
srihari@cedar.buffalo.edu

Topics

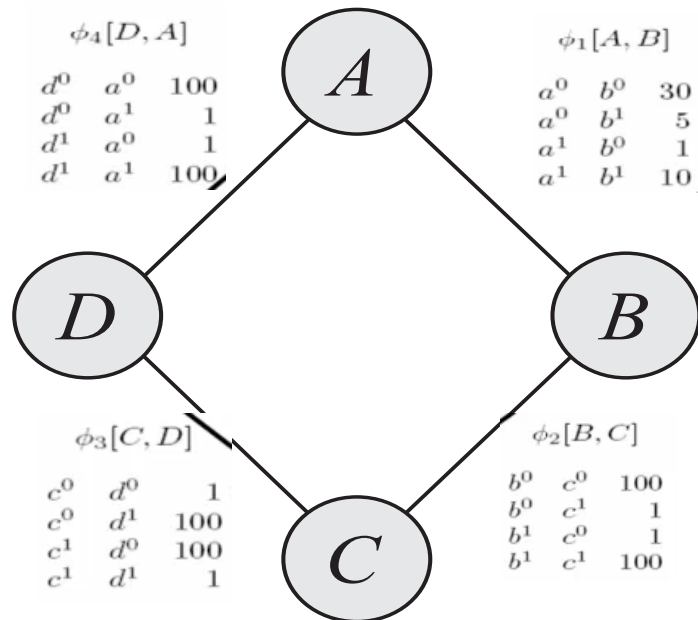
- The task: How to encode independencies in given distribution P in a graph structure G
- Theorems concerning
 - What type of Independencies?
 - Pairwise or Markov Blanket?
 - Relationship between P and G
 - I-map, Minimal I-Map or P-map?
 - Equivalent methods

From Distribution P to Graph \mathcal{H}

- How to encode independencies in given distribution P in a graph structure \mathcal{H} ?
- What sort of independencies to consider:
 - pairwise-local (intervening nodes blocked) or
 - local (Markov blanket)?
- Are we looking for an
 - I-map, i.e., $I(\mathcal{H}) \subseteq I(P)$
 - minimal I-map,
 - i.e., removal of an edge makes it not an I-map, or
 - Perfect map?

Global independencies: Separation

MN Graph with factors



$$P(a,b,c,d) = \frac{1}{Z} \phi_1(a,b) \cdot \phi_2(b,c) \cdot \phi_3(c,d) \cdot \phi_4(d,a)$$

where

$$Z = \sum_{a,b,c,d} \phi_1(a,b) \cdot \phi_2(b,c) \cdot \phi_3(c,d) \cdot \phi_4(d,a)$$

Global Independencies
are based on blockage of
influence

$$(B \perp D | A, C)$$

$$(A \perp C | B, D)$$

There are fewer
Local independencies
(Markov blankets)
to examine than global
independencies

Need for minimal I-map

- Notion of I-map is insufficient
 - Definition of I-map
 - I-map has $I(\mathcal{H}) \subseteq I(P)$
 - But complete graph (all nodes connected) encodes no independencies
- Thus we need notion of minimal I-map
 - Definition of Minimal I-map
 - Removal of an edge in \mathcal{H} renders it not an I-map

How to Construct Minimal I-map for Distribution P ?

- Consider two approaches
 1. Use pairwise Markov independencies
 - Pairs of nodes are independent given all others
 2. Use local independencies
 - Markov Blanket
- Both methods assume a *positive distribution*
 - For non-positive distributions neither pairwise nor local independencies imply global independencies
- We will see that methods are equivalent

Method 1: Distribution P to Graph \mathcal{H}

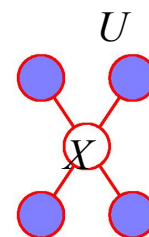
Use Pairwise Independencies

- Observe: If edge $\{X, Y\}$ is not in \mathcal{H} then X and Y must be independent given all other nodes, i.e., $(X \perp Y \mid \mathcal{X} - \{X, Y\})$
 - Conversely, If edge $\{X, Y\}$ is in \mathcal{H} then X and Y are dependent given other nodes, i.e., $(X \not\perp Y \mid \mathcal{X} - \{X, Y\})$
- Add edge between pair of nodes X, Y such that
 - $P \not\models (X \perp Y \mid \mathcal{X} - \{X, Y\})$ Symbol $\not\models$ means “does not support”
 - i.e., include an edge $X - Y$ for all pairs of nodes for which above equation holds, i.e., pairs that are dependent

Method 2: P to Graph \mathcal{H} with Local Independencies

- Use notion of Markov Blanket defined for P
- For each variable X , $MB_P(X)$ is a minimal set of nodes U satisfying definition of Markov Blanket
 - A set U is a Markov Blanket of X in distribution P if $X \notin U$ and if U is a minimal set of nodes such that

$$(X \perp \mathcal{X} - \{X\} - U \mid U) \in I(P)$$
 – i.e., minimal set of nodes that render X independent
- Define graph \mathcal{H} by introducing edge $\{X, Y\}$ for all $X, Y \in MB_P(X)$
 - Since X is dependent on nodes within its Markov Blanket



Equivalence of two methods

- Summarizing, let P be a positive distribution
- Thm 1 (pairwise):
 - Let H be defined by introducing edge $\{X, Y\}$ for all X, Y for which $P \not\models (X \perp Y | \chi - \{X, Y\})$ holds
 - Then H is a minimal I-map for P
- Thm 2 (local):
 - For each node X , let $MB_P(X)$ be a minimal set of nodes satisfying $(X \perp \chi - \{X\} - U | U) \in I(P)$
 - We define a graph H by introducing an edge $\{X, Y\}$ for all X and $Y \in MB_P(X)$
 - Then H is the unique minimal map for P

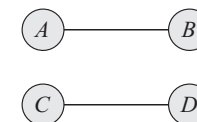
Example with non-positive distrib

- Equivalence does not hold when P not positive

– Ex: Non-positive distribution over binary variables A, B, C, D

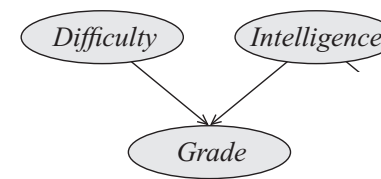
- Assigns non-zero probabilities only when they take on same value $P(a^1, b^1, c^1, d^1) = P(a^0, b^0, c^0, d^0) = 0.5$
- MB Algorithm
 - Since $P \models (A \perp C, D | B)$ $\{B\}$ is a legal choice for $MB_P(A)$
 - Similarly $P \models (C \perp A, B | D)$ $\{D\}$ is legal for $MB_P(C)$
 - Network satisfies all MB conditions
 - This network is not an I-map for the distribution
- Pairwise Algorithm
 - If we use pairwise I-map, network is empty
 - This network is also not an I-map for the distribution

A	B	C	D	P
0	0	0	0	0.5
1	0	0	0	0
1	1	0	0	0
1	1	1	0	0
0	0	0	1	0
0	0	1	1	0
0	1	1	1	0
1	1	1	1	0.5

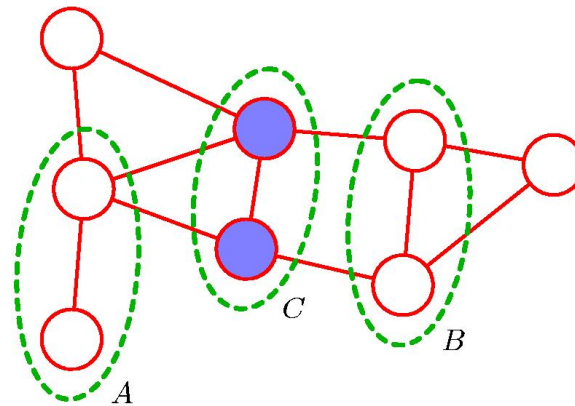


Does every distribution have a perfect map?

- Example: P is defined by BN with v-structure
- Let us try to represent it by a MN
 - We need edges $D--G$ and $I--G$
 - Can we omit edge between I and D ?
 - No, from D-separation, $(I \sim \perp D | G)$ which will be violated
 - Only minimal I-map is the fully connected graph
 - For removing an edge, gives an unwanted independence
 - Which does not capture $I \perp D$ which holds in P
 - Thus the minimal I-map is not a perfect map
 - i.e., $I(P) \neq I(H)$
- Our conclusion is given next



Global independence (Separation) and Positive Distributions



- Separation is sound and complete for determining conditional independence only for distributions for which \mathcal{H} is a perfect map