Undirected Graphical Models

Michael Gutmann

Probabilistic Modelling and Reasoning (INFR11134)
School of Informatics, University of Edinburgh

Spring Semester 2020

Recap

- We have seen that we can visualise pdfs/pmfs $p(\mathbf{x})$ without imposing an ordering or directionality of interaction between the random variables by using an undirected graph.
- The undirected graph allows us to read out independencies that must hold for $p(\mathbf{x})$.
- When we defined the graph for a pdf/pmf $p(\mathbf{x})$ the exact definition (e.g. numerical values) of $p(\mathbf{x})$ did not matter; we only used its factorisation.
- This enables us to define a set of probability distributions based on an undirected graph, i.e. an undirected graphical model.

Program

1. Definition of undirected graphical models
2. Independencies from graph separation
3. Further methods to determine independencies

Program

1. Definition of undirected graphical models
    - Via factorisation according to the graph
    - Maximal cliques
2. Independencies from graph separation
3. Further methods to determine independencies

Undirected graphical models

- We started with a pdf/pmf and associated a undirected graph with it.
- We now go the other way around and start with an undirected graph.
- *Definition* An undirected graphical model based on an undirected graph with d nodes and associated random variables x_i is the set of pdfs/pmfs that factorise as

$$p(x_1, \ldots, x_d) = \frac{1}{Z} \prod_c \phi_c(\mathcal{X}_c)$$

  where Z is the normalisation constant, $\phi_c(\mathcal{X}_c) \geq 0$, and the $\mathcal{X}_c$ correspond to the maximal cliques in the graph.
- $p(x_1, \ldots, x_d)$ as above are said to factorise over/according to the graph.

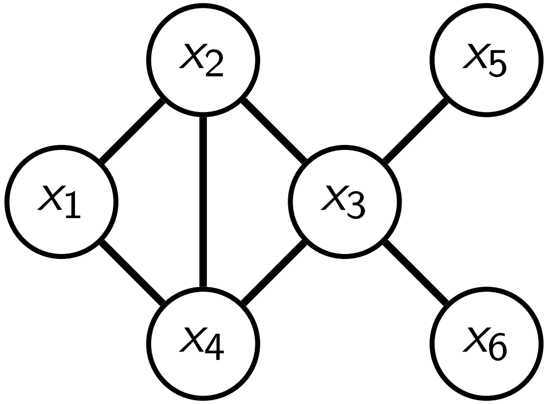
Remarks

- An undirected graph defines the pdfs/pmfs in form of Gibbs distributions.
- The undirected graphical model corresponds to a set of probability distributions. This is because we left the actual definition of the factors $\phi_c(\mathcal{X}_c)$ unspecified.
- People may also use "undirected graphical model" to refer to individual elements of the set (overloading of the name as for directed graphical models).
- Other names for an undirected graphical model: Markov network (MN), Markov random field (MRF)
- The $\mathcal{X}_c$ correspond to *maximal* cliques in the graph.

  Maximal clique: a set of fully connected nodes (clique) that is not contained in another clique.

Example

Undirected graph:



Random variables: $\mathbf{x} = (x_1, \ldots, x_6)$

Maximal cliques: $\{x_1, x_2, x_4\}$, $\{x_2, x_3, x_4\}$, $\{x_3, x_5\}$, $\{x_3, x_6\}$

Undirected graphical model: set of pdfs/pmfs $p(\mathbf{x})$ that factorise as:

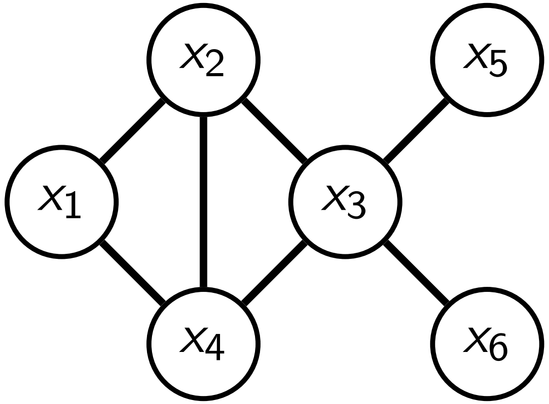
$$
\begin{aligned}
p(\mathbf{x}) &= \frac{1}{Z}\phi_1(x_1, x_2, x_4)\phi_2(x_2, x_3, x_4)\phi_3(x_3, x_5)\phi_4(x_3, x_6) \\
&\propto \phi_1(x_1, x_2, x_4)\phi_2(x_2, x_3, x_4)\phi_3(x_3, x_5)\phi_4(x_3, x_6)
\end{aligned}
$$

Why maximal cliques?

- The mapping from Gibbs distribution to graph is many to one. We may obtain the same graph for different Gibbs distributions, e.g.

$$p(\mathbf{x}) \propto \phi_1(x_1, x_2, x_4)\phi_2(x_2, x_3, x_4)\phi_3(x_3, x_5)\phi_4(x_3, x_6)$$

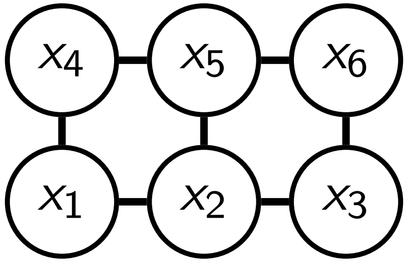
$$p(\mathbf{x}) \propto \tilde{\phi}_1(x_1, x_2)\tilde{\phi}_2(x_1, x_4)\tilde{\phi}_3(x_2, x_4)\tilde{\phi}_4(x_2, x_3)\tilde{\phi}_5(x_3, x_4)\tilde{\phi}_6(x_3, x_5)\tilde{\phi}_7(x_3, x_6)$$



- By using maximal cliques, we take a conservative approach and do not make additional assumptions on the factorisation.

Example (pairwise Markov network)

Graph:



Random variables: $\mathbf{x} = (x_1, \ldots, x_6)$

Maximal cliques: all neighbours

$\{x_1, x_2\} \quad \{x_2, x_3\} \quad \{x_4, x_5\} \quad \phi_6\{x_5, x_6\} \quad \{x_1, x_4\} \quad \{x_2, x_5\} \quad \phi_7\{x_3, x_6\}$

Undirected graphical model: set of pdfs/pmfs $p(\mathbf{x})$ that factorise as:

$$p(\mathbf{x}) \propto \phi_1(x_1, x_2)\phi_2(x_2, x_3)\phi_3(x_4, x_5)\phi_4(x_5, x_6)\phi_5(x_1, x_4)\phi_6(x_2, x_5)\phi_7(x_3, x_6)$$

Example of a pairwise Markov network.

Program

1. Definition of undirected graphical models
   - Via factorisation according to the graph
   - Maximal cliques

2. Independencies from graph separation

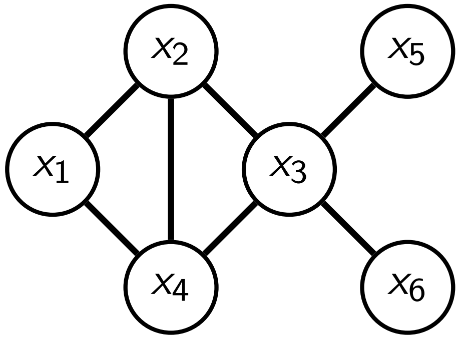
3. Further methods to determine independencies

Program

1. Definition of undirected graphical models
2. Independencies from graph separation
3. Further methods to determine independencies

Graph separation and conditional independence

- Undirected graph G:

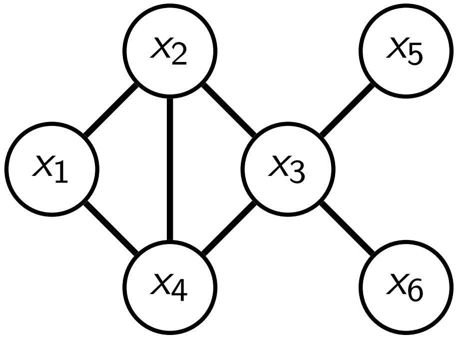


- The graph defines a set of pdfs/pmfs that factorise as $p(\mathbf{x}) \propto \phi_1(x_1, x_2, x_4)\phi_2(x_2, x_3, x_4)\phi_3(x_3, x_5)\phi_4(x_3, x_6)$
- Pick any *specific* instance p^* from the set.
- Visualising p^* as an undirected graph G^* gives the above graph or one with some edges removed.

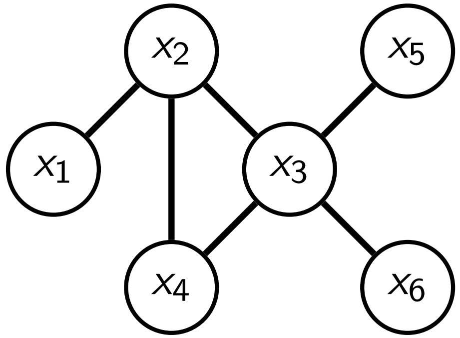
  Assume, for example, that p^* is such that the corresponding factor $\phi_1(x_1, x_2, x_4)$ does not depend on x_4. We would then not have an edge between x_1 and x_4.

Graph separation and conditional independence

G:



G^*:



- If a set Z separates some variables in G, it also separates them in G^*.
- Statistical independencies derived via graph separation using G must hold for p^* (but p^* may satisfy additional ones that we can't see using G)

$\Rightarrow$ p^* satisfies the global Markov property relative to G.

- This means that all pdfs/pmfs defined by an undirected graph satisfy the global Markov property relative to it.

Graph separation and conditional independence

Theorem:
Let G be the undirected graph and X, Y, Z three disjoint subsets of its nodes. If X and Y are separated by Z, then $X \perp\!\!\!\perp Y \mid Z$ for all probability distributions that factorise over the graph.

Important because:

1. the theorem allows us to read out (conditional) independencies from the undirected graph
2. no restriction on the sets X, Y, Z
3. the theorem shows that graph separation does not indicate false independence relations. ("Soundness" of the independence assertions.)

Graph separation and conditional independence

Theorem: If X and Y are not separated by Z in the graph then $X \not\perp\!\!\!\perp Y \mid Z$ in some probability distributions that factorise according to the graph.

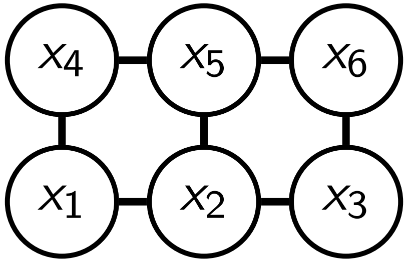
Optional, for those interested: A proof sketch can be found in Section 4.3.1.2 of *Probabilistic Graphical Models* by Koller and Friedman.

Remarks:

- The theorem implies that for some distributions, we may have $X \perp\!\!\!\perp Y \mid Z$ even though X and Y are not separated by Z. The separation criterion is not "complete" ("recall-rate" is not guaranteed to be 100%).
- This also means that the theorem only allows us to decide about independence and not about dependence.

Example (pairwise Markov network)

Graph:



Some independencies from global Markov property:

$$x_1, x_4 \perp\!\!\!\perp x_3, x_6 \mid x_2, x_5$$

$$x_1 \perp\!\!\!\perp \underbrace{x_5, x_6, x_3}_{\text{all } \setminus (x_1 \cup \mathrm{ne}_1)} \mid \underbrace{x_4, x_2}_{\mathrm{ne}_1} \qquad x_1 \perp\!\!\!\perp x_6 \mid \underbrace{x_2, x_3, x_4, x_5}_{\text{all without } x_1, x_6}$$

Last two are examples of the "local Markov property" and the "pairwise Markov property" relative to the undirected graph.

Program

1. Definition of undirected graphical models
2. Independencies from graph separation
3. Further methods to determine independencies

Program

1. Definition of undirected graphical models
2. Independencies from graph separation
3. Further methods to determine independencies
   - Local Markov property
   - Pairwise Markov property
   - Equivalence between factorisation and Markov properties for positive distributions
   - Markov blanket

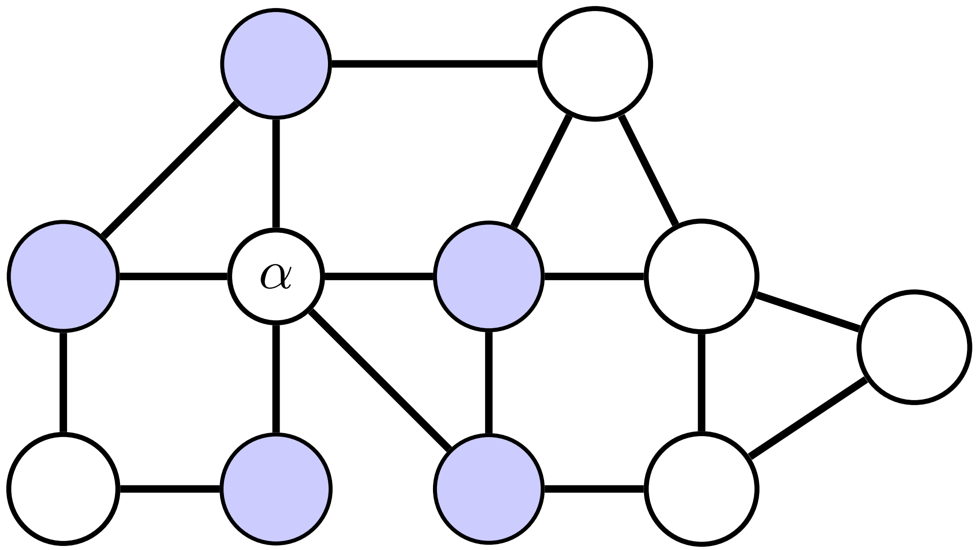
Local Markov property

Denote the set of all nodes by X and the neighbours of a node α by $\mathrm{ne}(\alpha)$.

- A probability distribution is said to satisfy the local Markov property relative to an undirected graph if

$$\alpha \perp\!\!\!\perp X \setminus (\alpha \cup \mathrm{ne}(\alpha)) \mid \mathrm{ne}(\alpha) \quad \text{for all nodes } \alpha \in X$$

- If p satisfies the global Markov property, then it satisfies the local Markov property. This is because $\mathrm{ne}(\alpha)$ blocks all trails to remaining nodes.



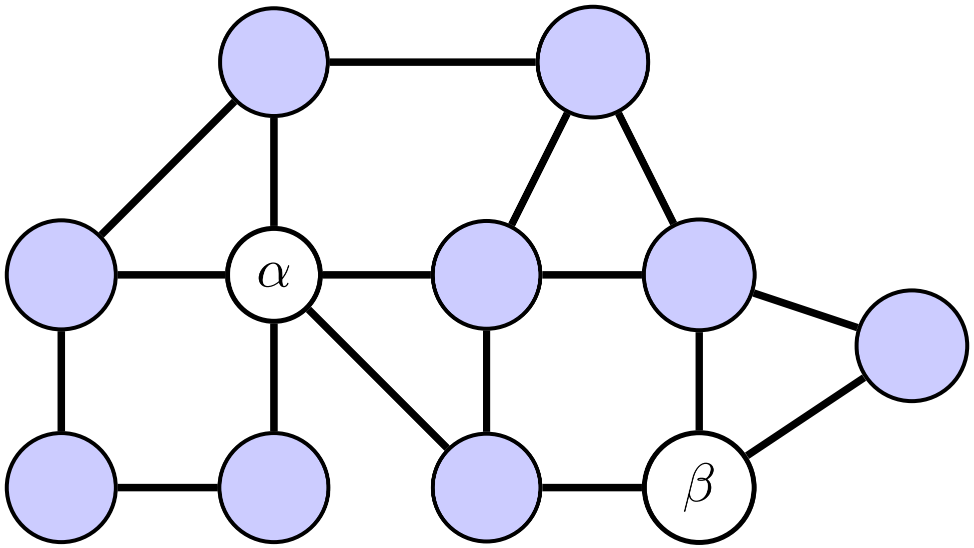
Pairwise Markov property

Denote the set of all nodes by X.

- A probability distribution is said to satisfy the pairwise Markov property relative to an undirected graph if

$$\alpha \perp\!\!\!\perp \beta \mid X \setminus \{\alpha, \beta\} \quad \text{for all non-neighbouring } \alpha, \beta \in X$$

- If p satisfies the local Markov property, then it satisfies the pairwise Markov property.



Summary

Let p be a pdf/pmf defined by the undirected graph G.

$$
\begin{gathered}
p \text{ factorises over } G \\
\Downarrow \\
p \text{ satisfies the global Markov property} \\
\Downarrow \\
p \text{ satisfies the local Markov property} \\
\Downarrow \\
p \text{ satisfies the pairwise Markov property}
\end{gathered}
$$

Do we have an equivalence?

- In directed graphical models, we had an equivalence of
  - factorisation,
  - ordered Markov property,
  - local directed Markov property, and
  - global directed Markov property.
- Do we have a similar equivalence for undirected graphical models?

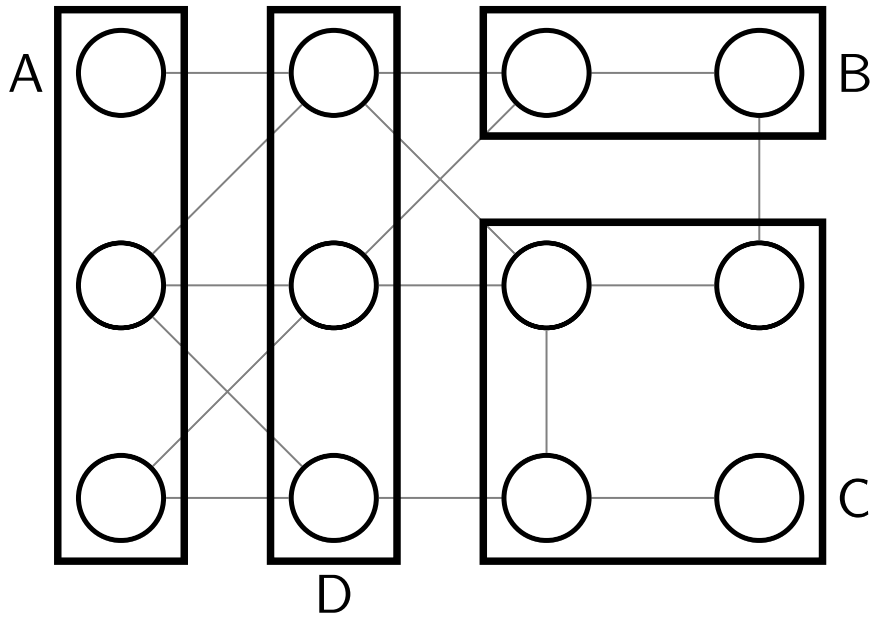
Yes, under some very mild condition

Intersection property

- The intersection property holds for all distributions with $p(\mathbf{x}) > 0$ for all values of $\mathbf{x}$ in its domain.

  (For a proof, see e.g. Lauritzen, 1996, Prop 3.1)
- Excludes deterministic relationships between the variables.
- Intersection property: Let A, B, C, D be sets of random variables

  If $A \perp\!\!\!\perp B \mid (C \cup D)$ and $A \perp\!\!\!\perp C \mid (B \cup D)$ then $A \perp\!\!\!\perp (B \cup C) \mid D$



From pairwise to global Markov property and factorisation

(For proofs, see e.g. Lauritzen, 1996, Section 3.2.)

- Let $p(x_1, \ldots, x_d)$ be a pdf/pmf that satisfies the intersection property for all disjoint subsets A, B, C, D of $\{x_1, \ldots, x_d\}$.
- If p satisfies the pairwise Markov property with respect to an undirected graph G then
  - p satisfies the global Markov property with respect to G, and
  - p factorises according to G.
- Hence: equivalence of factorisation and the global, local, and pairwise Markov properties for positive distributions.
- Equivalence known as Hammersely-Clifford theorem.
- Important e.g. for learning because prior knowledge may come in form of conditional independencies (the graph), which we can incorporate by working with Gibbs distributions that factorise accordingly.

Summary of the equivalences

Given a undirected graph with nodes (random variables) x_i and maximal cliques $\mathcal{X}_c$, we have the following equivalences:

$p(\mathbf{x})$ factorises over the graph		$p(x_1, \ldots, x_d) = \frac{1}{Z} \prod_c \phi_c(\mathcal{X}_c), \quad \phi_c(\mathcal{X}_c) > 0$
	$\Updownarrow$	
$p(\mathbf{x})$ satisfies the pairwise MP		$\alpha \perp\!\!\!\perp \beta \mid \{x_1, \ldots, x_d\} \setminus \{\alpha, \beta\}$
	$\Updownarrow$	
$p(\mathbf{x})$ satisfies the local MP		$\alpha \perp\!\!\!\perp \{x_1, \ldots, x_d\} \setminus (\alpha \cup \mathrm{ne}(\alpha)) \mid \mathrm{ne}(\alpha)$
	$\Updownarrow$	
$p(\mathbf{x})$ satisfies the global MP		all independencies asserted by graph separation

(MP: Markov property)

Broadly speaking, the graph serves two related purposes:

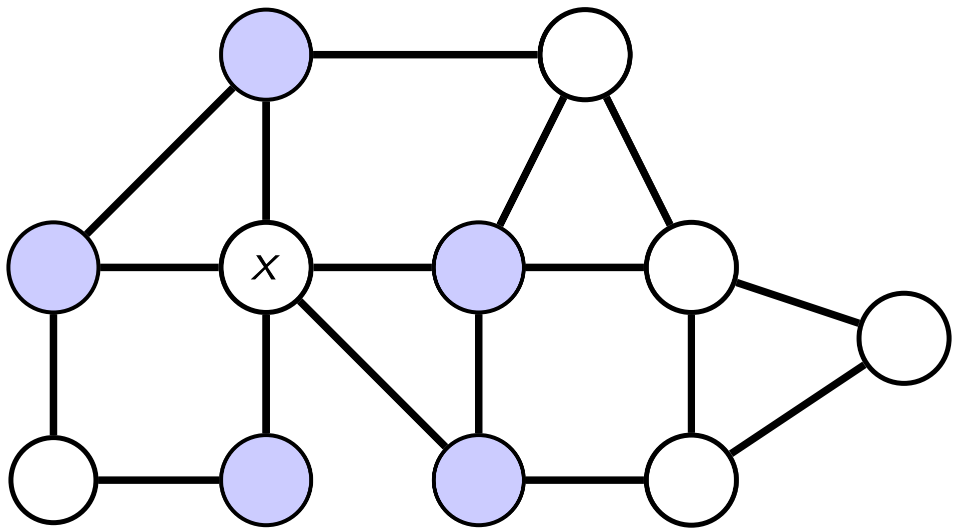
1. it tells us how distributions factorise
2. it represents the independence assumptions made

Markov blanket

What is the minimal set of variables such that knowing their values makes x independent from the rest?

From local Markov property: $\mathrm{MB}(x) = \mathrm{ne}(x)$:

$$x \perp\!\!\!\perp \{\text{all variables} \setminus (x \cup \mathrm{ne}(x))\} \mid \mathrm{ne}(x)$$



Program recap

1. Definition of undirected graphical models
   - Via factorisation according to the graph
   - Maximal cliques

2. Independencies from graph separation

3. Further methods to determine independencies
   - Local Markov property
   - Pairwise Markov property
   - Equivalence between factorisation and Markov properties for positive distributions
   - Markov blanket