Independencies and Undirected Graphs

Michael Gutmann

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

Spring semester 2019

Recap

- The number of free parameters in probabilistic models increases with the number of random variables involved.
- Making statistical independence assumptions reduces the number of free parameters that need to be specified.
- Starting with the chain rule and an ordering of the random variables, we used statistical independencies to simplify the representation.
- We thus obtained a factorisation in terms of a product of conditional pdfs that we visualised as a DAG.
- In turn, we used DAGs to define sets of distributions ("directed graphical models").
- We discussed independence properties satisfied by the distributions, d-separation, and the equivalence to the factorisation.

The directionality in directed graphical models

- So far we mainly exploited the property

$$\mathbf{x} \perp\!\!\!\perp \mathbf{y} \mid \mathbf{z} \iff p(\mathbf{y}|\mathbf{x}, \mathbf{z}) = p(\mathbf{y}|\mathbf{z})$$

- But when working with $p(\mathbf{y}|\mathbf{x}, \mathbf{z})$ we impose an ordering or directionality from $\mathbf{x}$ and $\mathbf{z}$ to $\mathbf{y}$.
- Directionality matters in directed graphical models

$$x \longrightarrow z \longrightarrow y \qquad \text{versus} \qquad x \longrightarrow z \longleftarrow y$$

- In some cases, directionality is natural but in others we do not want to choose one direction over another.
- We now discuss how to represent probability distributions and independencies in a symmetric manner without assuming a directionality or ordering of the variables.

Program

1. Representing probability distributions without imposing a directionality between the random variables

2. Separation in undirected graphs and statistical independencies

Program

1. Representing probability distributions without imposing a directionality between the random variables
   - Factorisation and statistical independence
   - Gibbs distributions
   - Visualising Gibbs distributions with undirected graphs
   - Conditioning corresponds to removing nodes and edges from the graph

2. Separation in undirected graphs and statistical independencies

Further characterisation of statistical independence

- From tutorials: For non-negative functions $a(\mathbf{x}, \mathbf{z}), b(\mathbf{y}, \mathbf{z})$:

$$\mathbf{x} \perp\!\!\!\perp \mathbf{y} \mid \mathbf{z} \Longleftrightarrow p(\mathbf{x}, \mathbf{y}, \mathbf{z}) = a(\mathbf{x}, \mathbf{z}) b(\mathbf{y}, \mathbf{z})$$

- More general version of $p(\mathbf{x}, \mathbf{y}, \mathbf{z}) = p(\mathbf{x}|\mathbf{z}) p(\mathbf{y}|\mathbf{z}) p(\mathbf{z})$
- No directionality or ordering of the variables is imposed.
- Unconditional version: For non-negative functions $a(\mathbf{x}), b(\mathbf{y})$:

$$\mathbf{x} \perp\!\!\!\perp \mathbf{y} \Longleftrightarrow p(\mathbf{x}, \mathbf{y}) = a(\mathbf{x}) b(\mathbf{y})$$

- The important point is the factorisation of $p(\mathbf{x}, \mathbf{y}, \mathbf{z})$ into two factors:
  - if the factors share a variable $\mathbf{z}$, then we have conditional independence,
  - if not, we have unconditional independence.

Further characterisation of statistical independence

- Since $p(\mathbf{x}, \mathbf{y}, \mathbf{z})$ must sum (integrate) to one, we must have

$$\sum_{\mathbf{x},\mathbf{y},\mathbf{z}} a(\mathbf{x}, \mathbf{z}) b(\mathbf{y}, \mathbf{z}) = 1$$

- Normalisation condition often ensured by re-defining $a(\mathbf{x}, \mathbf{z}) b(\mathbf{y}, \mathbf{z})$:

$$p(\mathbf{x}, \mathbf{y}, \mathbf{z}) = \frac{1}{Z} \phi_A(\mathbf{x}, \mathbf{z}) \phi_B(\mathbf{y}, \mathbf{z}) \qquad Z = \sum_{\mathbf{x},\mathbf{y},\mathbf{z}} \phi_A(\mathbf{x}, \mathbf{z}) \phi_B(\mathbf{y}, \mathbf{z})$$

- Z: normalisation constant (related to partition function, see later)
- ϕ_i: factors (also called potential functions).
  Do generally not correspond to (conditional) probabilities.
  They measure "compatibility", "agreement", or "affinity"

What does it mean?

$$\mathbf{x} \perp\!\!\!\perp \mathbf{y} \mid \mathbf{z} \Longleftrightarrow p(\mathbf{x}, \mathbf{y}, \mathbf{z}) = \frac{1}{Z}\phi_A(\mathbf{x}, \mathbf{z})\phi_B(\mathbf{y}, \mathbf{z})$$

"$\Rightarrow$" If we want our model to satisfy $\mathbf{x} \perp\!\!\!\perp \mathbf{y} \mid \mathbf{z}$ we should write the pdf (pmf) as

$$p(\mathbf{x}, \mathbf{y}, \mathbf{z}) \propto \phi_A(\mathbf{x}, \mathbf{z})\phi_B(\mathbf{y}, \mathbf{z})$$

"$\Leftarrow$" If the pdf (pmf) can be written as $p(\mathbf{x}, \mathbf{y}, \mathbf{z}) \propto \phi_A(\mathbf{x}, \mathbf{z})\phi_B(\mathbf{y}, \mathbf{z})$ then we have $\mathbf{x} \perp\!\!\!\perp \mathbf{y} \mid \mathbf{z}$

equivalent for unconditional version

Example

Consider $p(x_1, x_2, x_3, x_4) \propto \phi_1(x_1, x_2)\phi_2(x_2, x_3)\phi_3(x_4)$

What independencies does p satisfy?

- We can write

$$p(x_1, x_2, x_3, x_4) \propto \underbrace{[\phi_1(x_1, x_2)\phi_2(x_2, x_3)]}_{\tilde{\phi}_1(x_1,x_2,x_3)}[\phi_3(x_4)]$$
$$\propto \tilde{\phi}_1(x_1, x_2, x_3)\phi_3(x_4)$$

  so that $x_4 \perp\!\!\!\perp x_1, x_2, x_3$.

- Integrating out x_4 gives

$$p(x_1, x_2, x_3) = \int p(x_1, x_2, x_3, x_4)\mathrm{d}x_4 \propto \phi_1(x_1, x_2)\phi_2(x_2, x_3)$$

  so that $x_1 \perp\!\!\!\perp x_3 \mid x_2$

Gibbs distributions

- Example is a special case of a class of pdfs/pmfs that factorise as
$$p(x_1, \ldots, x_d) = \frac{1}{Z} \prod_c \phi_c(\mathcal{X}_c)$$
  - $\mathcal{X}_c \subseteq \{x_1, \ldots, x_d\}$
  - ϕ_c are non-negative factors (potential functions)
    Do generally not correspond to (conditional) probabilities.
    They measure "compatibility", "agreement", or "affinity"
  - Z is a normalising constant so that $p(x_1, \ldots, x_d)$ integrates (sums) to one.
- Known as Gibbs (or Boltzmann) distributions
- $\tilde{p}(x_1, \ldots, x_d) = \prod_c \phi_c(\mathcal{X}_c)$ is an example of an unnormalised model: $\tilde{p} \geq 0$ but does not necessarily integrate (sum) to one.

Energy-based model

- With $\phi_c(\mathcal{X}_c) = \exp\left(-E_c(\mathcal{X}_c)\right)$, we have equivalently

$$p(x_1, \ldots, x_d) = \frac{1}{Z} \exp\left[-\sum_c E_c(\mathcal{X}_c)\right]$$

- $\sum_c E_c(\mathcal{X}_c)$ is the energy of the configuration $(x_1, \ldots, x_d)$.
  low energy $\iff$ high probability

Example

Other examples of Gibbs distributions:

$$
\begin{aligned}
p(x_1, \ldots, x_6) &\propto \phi_1(x_1, x_2, x_4)\phi_2(x_2, x_3, x_4)\phi_3(x_3, x_5)\phi_4(x_3, x_6) \\
p(x_1, \ldots, x_6) &\propto \phi_1(x_1, x_2)\phi_2(x_2, x_3)\phi_3(x_2, x_5)\phi_4(x_1, x_4)\phi_5(x_4, x_5) \\
&\qquad \phi_6(x_5, x_6)\phi_7(x_3, x_6)?
\end{aligned}
$$

Independencies?

- In principle, the independencies follow from

$$\mathbf{x} \perp\!\!\!\perp \mathbf{y} \mid \mathbf{z} \Longleftrightarrow p(\mathbf{x}, \mathbf{y}, \mathbf{z}) \propto \phi_A(\mathbf{x}, \mathbf{z})\phi_B(\mathbf{y}, \mathbf{z})$$

  with appropriately defined factors ϕ_A and ϕ_B.
- But the mathematical manipulations of grouping together factors and integrating variables out become unwieldy.

Let us use graphs to better see what's going on.

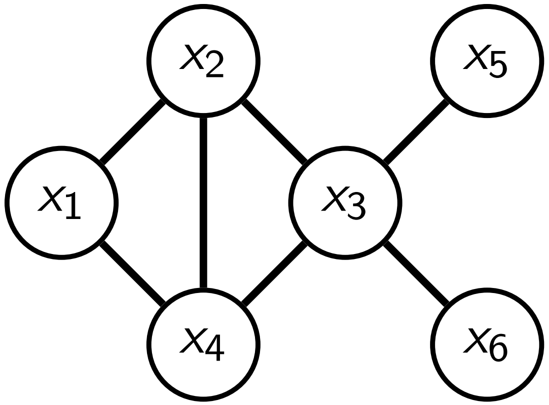
Visualising Gibbs distributions with undirected graphs

$p(x_1, \ldots, x_d) \propto \prod_c \phi_c(\mathcal{X}_c)$

- Node for each x_i
- For all factors ϕ_c: draw an undirected edge between all x_i and x_j that belong to $\mathcal{X}_c$
- Results in a fully-connected subgraph for all x_i that are part of the same factor (this subgraph is called a clique).

Example:

Graph for $p(x_1, \ldots, x_6) \propto \phi_1(x_1, x_2, x_4)\phi_2(x_2, x_3, x_4)\phi_3(x_3, x_5)\phi_4(x_3, x_6)$



Effect of conditioning

Let $p(x_1, \ldots, x_6) \propto \phi_1(x_1, x_2, x_4)\phi_2(x_2, x_3, x_4)\phi_3(x_3, x_5)\phi_4(x_3, x_6)$.

- What is $p(x_1, x_2, x_4, x_5, x_6 | x_3 = \alpha)$?
- By definition $p(x_1, x_2, x_4, x_5, x_6 | x_3 = \alpha)$

$$
\begin{aligned}
&= \frac{p(x_1, x_2, x_3 = \alpha, x_4, x_5, x_6)}{\int p(x_1, x_2, x_3 = \alpha, x_4, x_5, x_6) \mathrm{d}x_1 \mathrm{d}x_2 \mathrm{d}x_4 \mathrm{d}x_5 \mathrm{d}x_6} \\
&= \frac{\phi_1(x_1, x_2, x_4)\phi_2(x_2, \alpha, x_4)\phi_3(\alpha, x_5)\phi_4(\alpha, x_6)}{\int \phi_1(x_1, x_2, x_4)\phi_2(x_2, \alpha, x_4)\phi_3(\alpha, x_5)\phi_4(\alpha, x_6) \mathrm{d}x_1 \mathrm{d}x_2 \mathrm{d}x_4 \mathrm{d}x_5 \mathrm{d}x_6} \\
&= \frac{1}{Z(\alpha)} \phi_1(x_1, x_2, x_4)\phi_2^\alpha(x_2, x_4)\phi_3^\alpha(x_5)\phi_4^\alpha(x_6)
\end{aligned}
$$

- Gibbs distribution with derived factors ϕ_i^α of reduced domain and new normalisation "constant" $Z(\alpha)$
- Note that $Z(\alpha)$ depends on the conditioning value α.

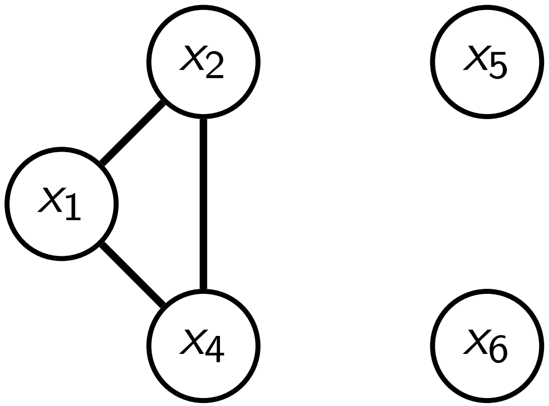
Effect of conditioning

Let $p(x_1, \ldots, x_6) \propto \phi_1(x_1, x_2, x_4)\phi_2(x_2, x_3, x_4)\phi_3(x_3, x_5)\phi_4(x_3, x_6)$.

- Conditional $p(x_1, x_2, x_4, x_5, x_6 | x_3 = \alpha)$ is

$$\frac{1}{Z(\alpha)}\phi_1(x_1, x_2, x_4)\phi_2^\alpha(x_2, x_4)\phi_3^\alpha(x_5)\phi_4^\alpha(x_6)$$

- Conditioning on variables removes the corresponding nodes and connecting edges from the undirected graph



Program

1. Representing probability distributions without imposing a directionality between the random variables
   - Factorisation and statistical independence
   - Gibbs distributions
   - Visualising Gibbs distributions with undirected graphs
   - Conditioning corresponds to removing nodes and edges from the graph

2. Separation in undirected graphs and statistical independencies

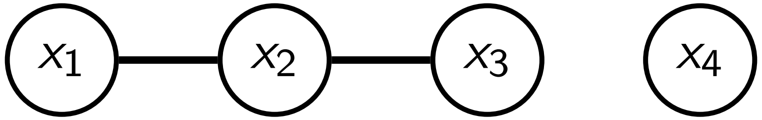
Program

1. Representing probability distributions without imposing a directionality between the random variables

2. Separation in undirected graphs and statistical independencies
   - Separation in undirected graphs
   - Statistical independencies from graph separation
   - Global Markov property

Relating graph properties to independencies

- Consider $p(x_1, x_2, x_3, x_4) \propto \phi_1(x_1, x_2)\phi_2(x_2, x_3)\phi_3(x_4)$ from before
- We have seen:
  - $x_4 \perp\!\!\!\perp x_1, x_2, x_3$
  - $x_1 \perp\!\!\!\perp x_3 \mid x_2$
- Graph:

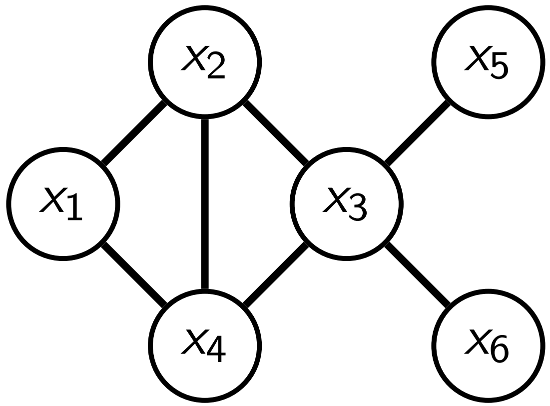


- In the graph, x_4 is separated from x_1, x_2, x_3.

  Starting at x_4, we cannot reach x_1, x_2, or x_3 (and vice versa). In other words, all trails from x_4 to x_1, x_2, x_3 are "blocked".
- In the graph, x_1 and x_3 are separated by x_2. In other words, all trails from x_1 to x_3 are blocked by x_2 (when removing x_2 from the graph, we cannot reach x_3 from x_1 and vice versa)

Relating graph properties to independencies

- Example:
  $p(x_1, \ldots, x_6) \propto \phi_1(x_1, x_2, x_4)\phi_2(x_2, x_3, x_4)\phi_3(x_3, x_5)\phi_4(x_3, x_6)$
- Graph:



- x_3 separates $\{x_1, x_2, x_4\}$ and $\{x_5, x_6\}$
  In other words, x_3 blocks all trails from $\{x_1, x_2, x_4\}$ to $\{x_5, x_6\}$
- Do we have $x_1, x_2, x_4 \perp\!\!\!\perp x_5, x_6 \mid x_3$?

Relating graph properties to independencies

$$p(x_1, \ldots, x_6) \propto \phi_1(x_1, x_2, x_4)\phi_2(x_2, x_3, x_4)\phi_3(x_3, x_5)\phi_4(x_3, x_6)$$

- Do we have $x_1, x_2, x_4 \perp\!\!\!\perp x_5, x_6 \mid x_3$?
- Group the factors

$$p(x_1, \ldots, x_6) \propto \underbrace{\phi_1(x_1, x_2, x_4)\phi_2(x_2, x_3, x_4)}_{\phi_A(x_1, x_2, x_4, x_3)} \underbrace{\phi_3(x_3, x_5)\phi_4(x_3, x_6)}_{\phi_B(x_5, x_6, x_3)}$$

- Takes the form

$$p(\mathbf{x}, \mathbf{y}, \mathbf{z}) \propto \phi_A(\mathbf{x}, \mathbf{z})\phi_B(\mathbf{y}, \mathbf{z})$$

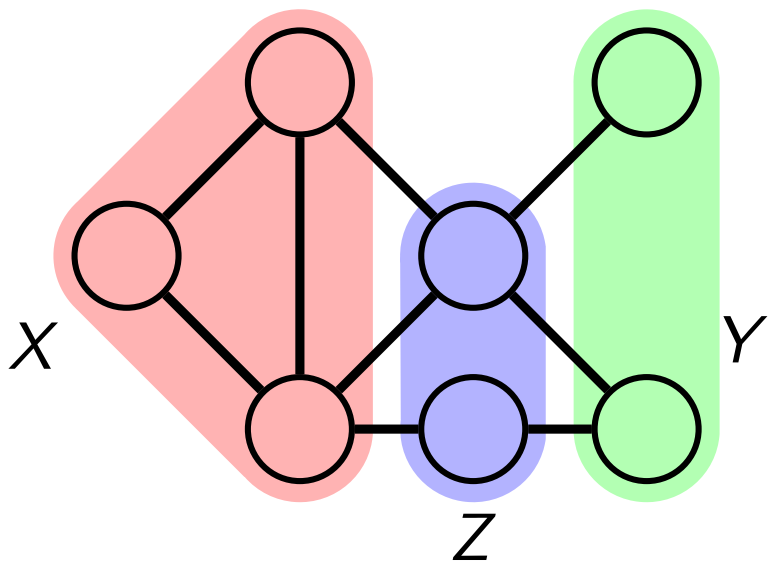
with $\mathbf{x} = (x_1, x_2, x_4)$, $\mathbf{y} = (x_5, x_6)$, $\mathbf{z} = x_3$

- Hence: $x_1, x_2, x_4 \perp\!\!\!\perp x_5, x_6 \mid x_3$ holds indeed.

Separation in undirected graphs

Let X, Y, Z be three disjoint set of nodes in an undirected graph.

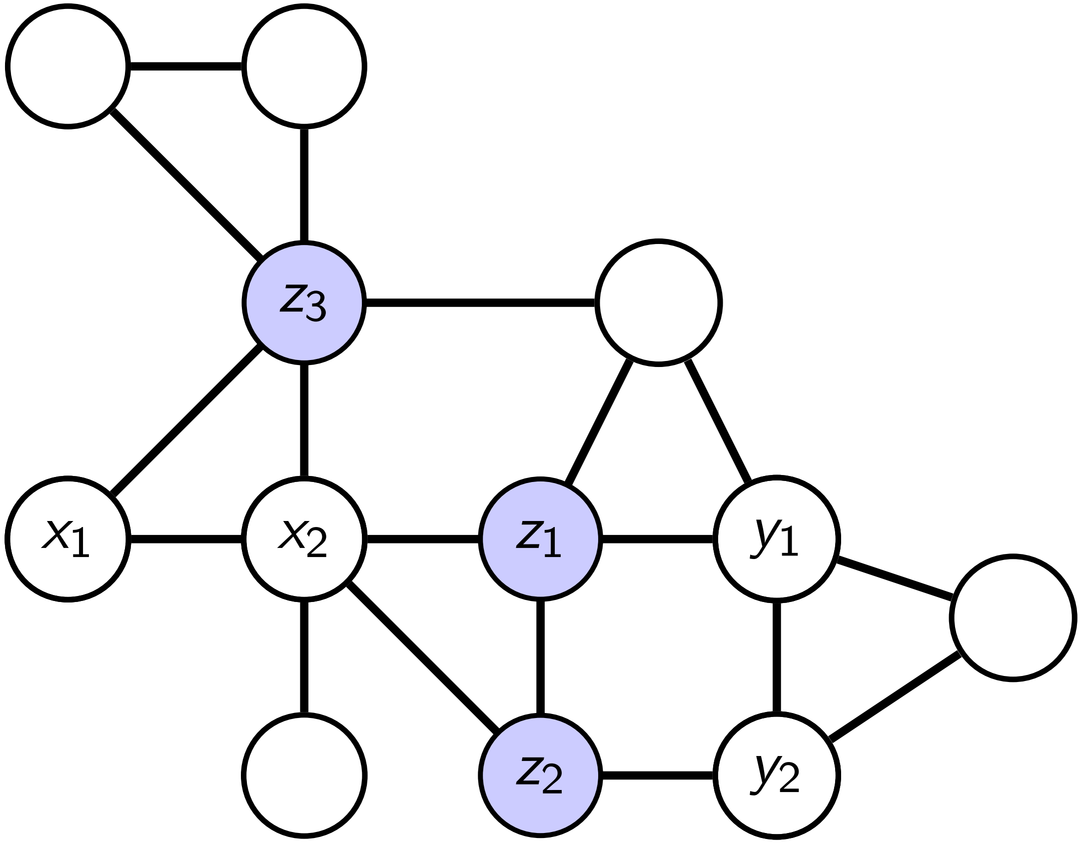
- X and Y are separated by Z if every trail from any node in X to any node in Y passes through at least one node of Z.
- In other words:
  - all trails from X to Y are blocked by Z
  - removing Z from the graph leaves X and Y disconnected.
  - Nodes are valves; open by default but closed when part of Z.



Statistical independencies from graph separation

Assume $p(x_1, \ldots, x_d) \propto \prod_c \phi_c(\mathcal{X}_c)$, with $\mathcal{X}_c \subset \{x_1, \ldots, x_d\}$ can be visualised as the graph below.

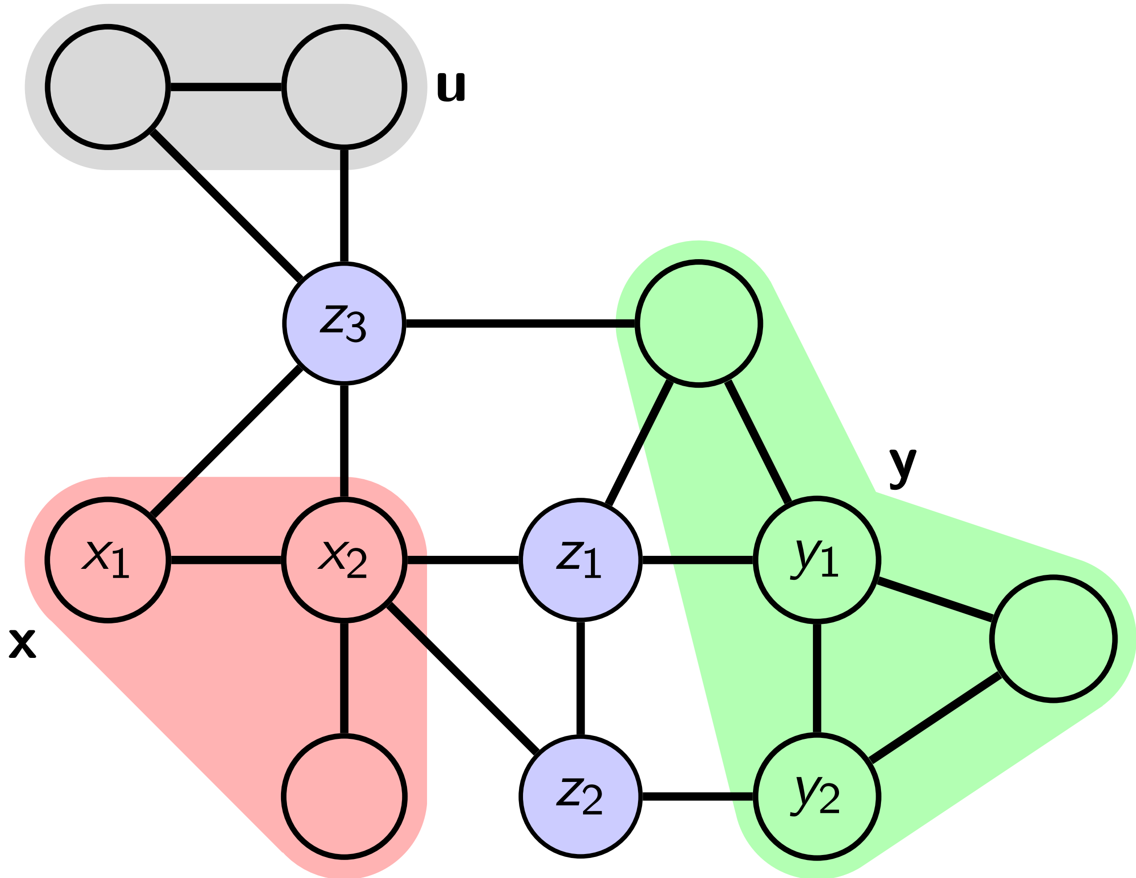
Do we have $x_1, x_2 \perp\!\!\!\perp y_1, y_2 \mid z_1, z_2, z_3$?



Statistical independencies from graph separation

Assume $p(x_1, \ldots, x_d) \propto \prod_c \phi_c(\mathcal{X}_c)$, with $\mathcal{X}_c \subset \{x_1, \ldots, x_d\}$ can be visualised as the graph below.

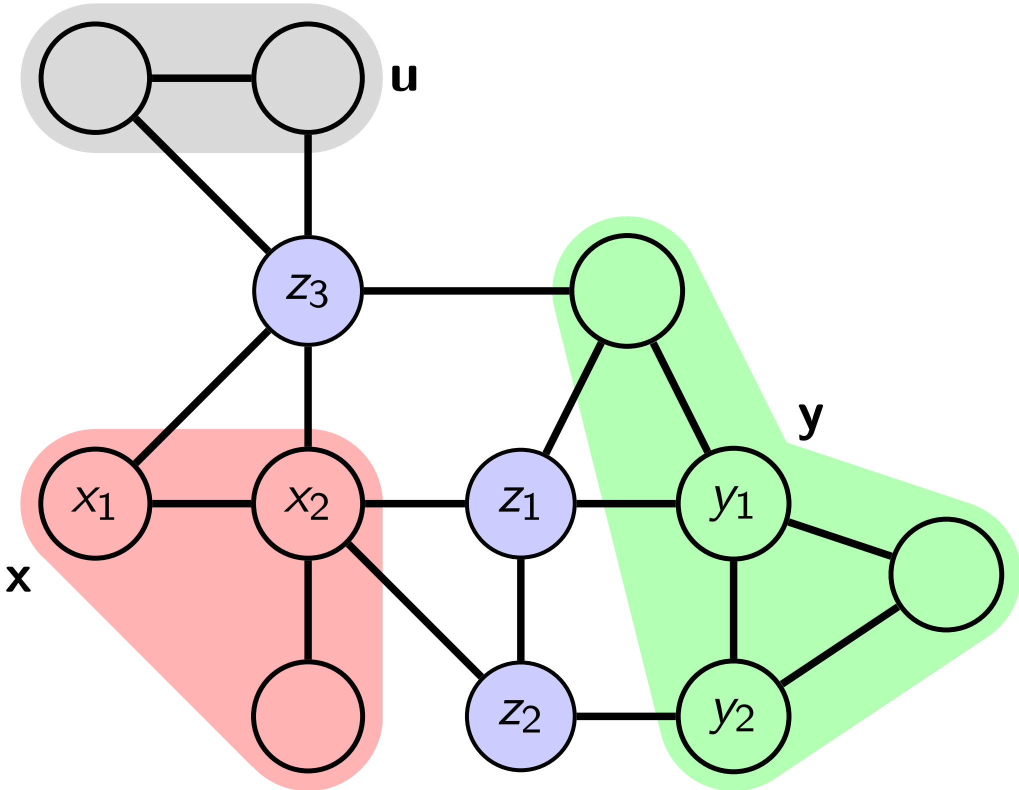
Do we have $\mathbf{x} \perp\!\!\!\perp \mathbf{y} \mid z_1, z_2, z_3$?



Statistical independencies from graph separation

- With $\mathbf{z} = (z_1, z_2, z_3)$, all x_i belong to one of the $\mathbf{x}, \mathbf{y}, \mathbf{z}$, or $\mathbf{u}$.
- We thus have $p(x_1, \ldots, x_d) = p(\mathbf{x}, \mathbf{y}, \mathbf{z}, \mathbf{u})$ and we can group the factors ϕ_c together so that

$$p(\mathbf{x}, \mathbf{y}, \mathbf{z}, \mathbf{u}) \propto \phi_1(\mathbf{x}, \mathbf{z})\phi_2(\mathbf{y}, \mathbf{z})\phi_3(\mathbf{u}, \mathbf{z})$$



Statistical independencies from graph separation

- Integrating (summing) out $\mathbf{u}$ gives

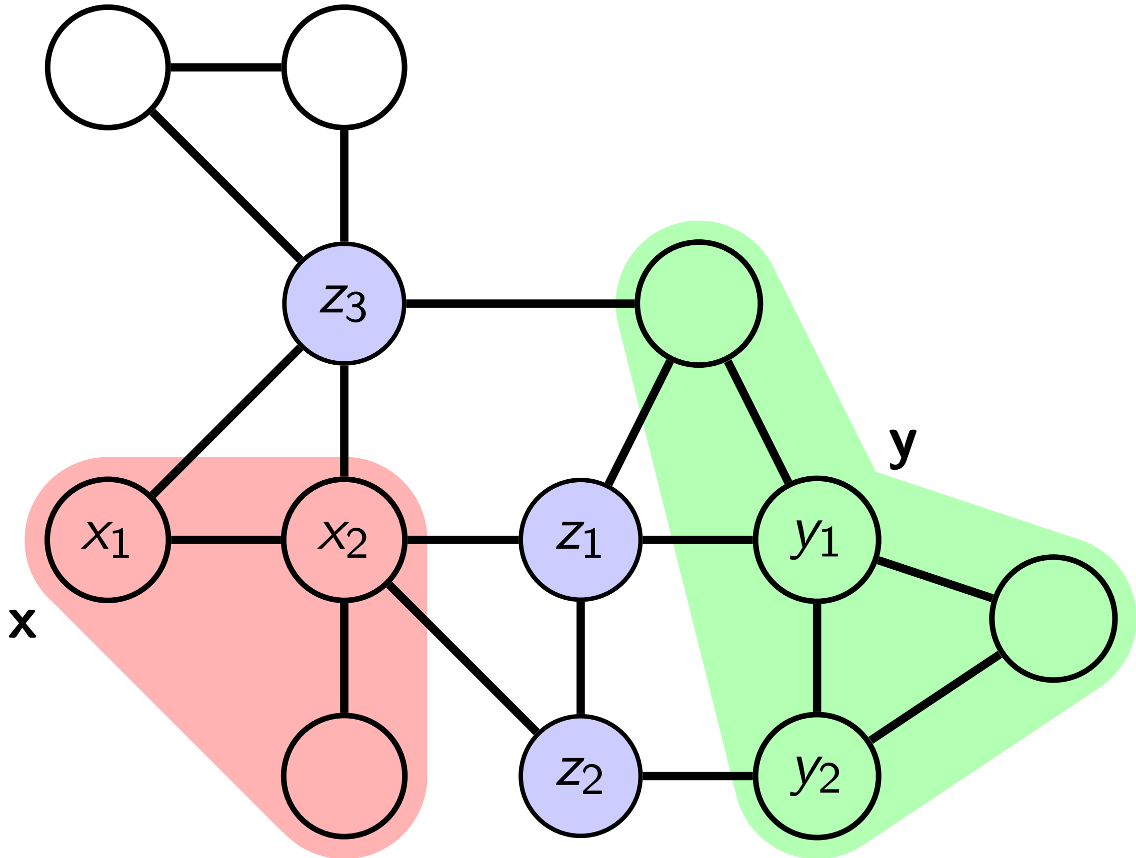
$$
\begin{aligned}
p(\mathbf{x},\mathbf{y},\mathbf{z}) &= \sum_{\mathbf{u}} p(\mathbf{x},\mathbf{y},\mathbf{z},\mathbf{u}) && (1)\\
&\propto \sum_{\mathbf{u}} \phi_1(\mathbf{x},\mathbf{z})\phi_2(\mathbf{y},\mathbf{z})\phi_3(\mathbf{u},\mathbf{z}) && (2)\\
\text{(distributive law)} \quad &\propto \phi_1(\mathbf{x},\mathbf{z})\phi_2(\mathbf{y},\mathbf{z})\sum_{\mathbf{u}} \phi_3(\mathbf{u},\mathbf{z}) && (3)\\
&\propto \phi_1(\mathbf{x},\mathbf{z})\phi_2(\mathbf{y},\mathbf{z})\tilde{\phi}(\mathbf{z}) && (4)\\
&\propto \phi_A(\mathbf{x},\mathbf{z})\phi_B(\mathbf{y},\mathbf{z}) && (5)
\end{aligned}
$$

- And $p(\mathbf{x},\mathbf{y},\mathbf{z}) \propto \phi_A(\mathbf{x},\mathbf{z})\phi_B(\mathbf{y},\mathbf{z})$ means $\mathbf{x} \perp\!\!\!\perp \mathbf{y} \mid \mathbf{z}$

Statistical independencies from graph separation

Assume $p(x_1, \ldots, x_d) \propto \prod_c \phi_c(\mathcal{X}_c)$, with $\mathcal{X}_c \subset \{x_1, \ldots, x_d\}$ can be visualised as the graph below.

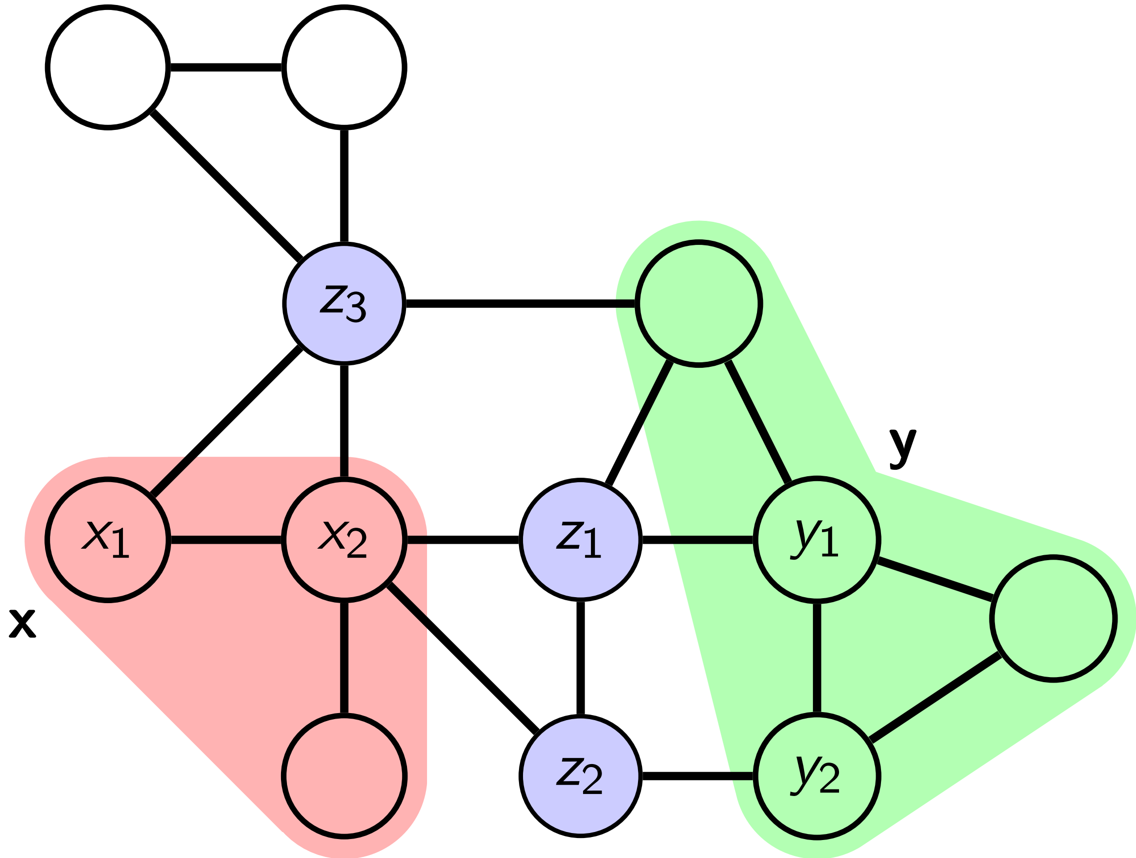
We have shown that if $\mathbf{x}$ and $\mathbf{y}$ are separated by $\mathbf{z}$, then $\mathbf{x} \perp\!\!\!\perp \mathbf{y} \mid \mathbf{z}$.



Statistical independencies from graph separation

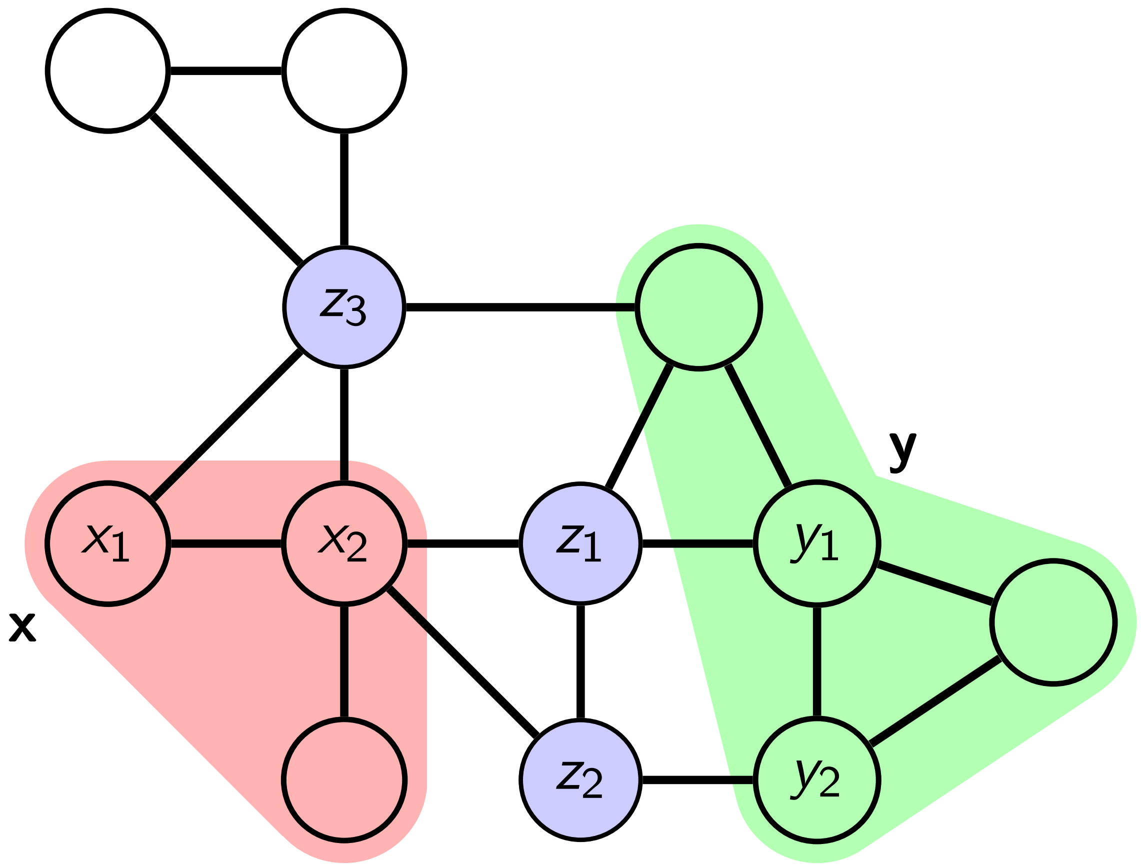
Assume $p(x_1, \ldots, x_d) \propto \prod_c \phi_c(\mathcal{X}_c)$, with $\mathcal{X}_c \subset \{x_1, \ldots, x_d\}$ can be visualised as the graph below.

So do we have $x_1, x_2 \perp\!\!\!\perp y_1, y_2 \mid z_1, z_2, z_3$?



Statistical independencies from graph separation

- From tutorial: $x \perp\!\!\!\perp \{y, w\} \mid z$ implies $x \perp\!\!\!\perp y \mid z$
- Hence $\mathbf{x} \perp\!\!\!\perp \mathbf{y} \mid z_1, z_2, z_3$ implies $x_1, x_2 \perp\!\!\!\perp y_1, y_2 \mid z_1, z_2, z_3$.



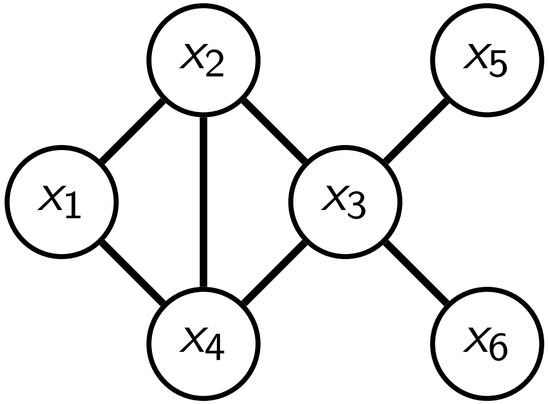
Summary

Theorem:
Let G be the undirected graph for $p(x_1, \ldots, x_d) \propto \prod_c \phi_c(\mathcal{X}_c)$, and X, Y, Z three disjoint subsets of $\{x_1, \ldots, x_d\}$. If X and Y are separated by Z in G, then p is such that $X \perp\!\!\!\perp Y \mid Z$.

- Important because:
  1. the theorem allows us to read out (conditional) independencies from the undirected graph
  2. the theorem shows that graph separation does not indicate false independence relations. ("Soundness" of the independence assertions.)
- We say that $p(x_1, \ldots, x_d)$ satisfies the global Markov property relative to G.

Example

- $p(x_1, \dots, x_6) \propto \phi_1(x_1, x_2, x_4)\phi_2(x_2, x_3, x_4)\phi_3(x_3, x_5)\phi_4(x_3, x_6)$
- Graph



- Example independencies:
  $x_1 \perp\!\!\!\perp \{x_3, x_5, x_6\} \mid x_2, x_4$ $\quad$ $x_2 \perp\!\!\!\perp x_6 \mid x_3$ $\quad$ $x_5 \perp\!\!\!\perp x_6 \mid x_3$

Program recap

1. Representing probability distributions without imposing a directionality between the random variables
    - Factorisation and statistical independence
    - Gibbs distributions
    - Visualising Gibbs distributions with undirected graphs
    - Conditioning corresponds to removing nodes and edges from the graph

2. Separation in undirected graphs and statistical independencies
    - Separation in undirected graphs
    - Statistical independencies from graph separation
    - Global Markov property