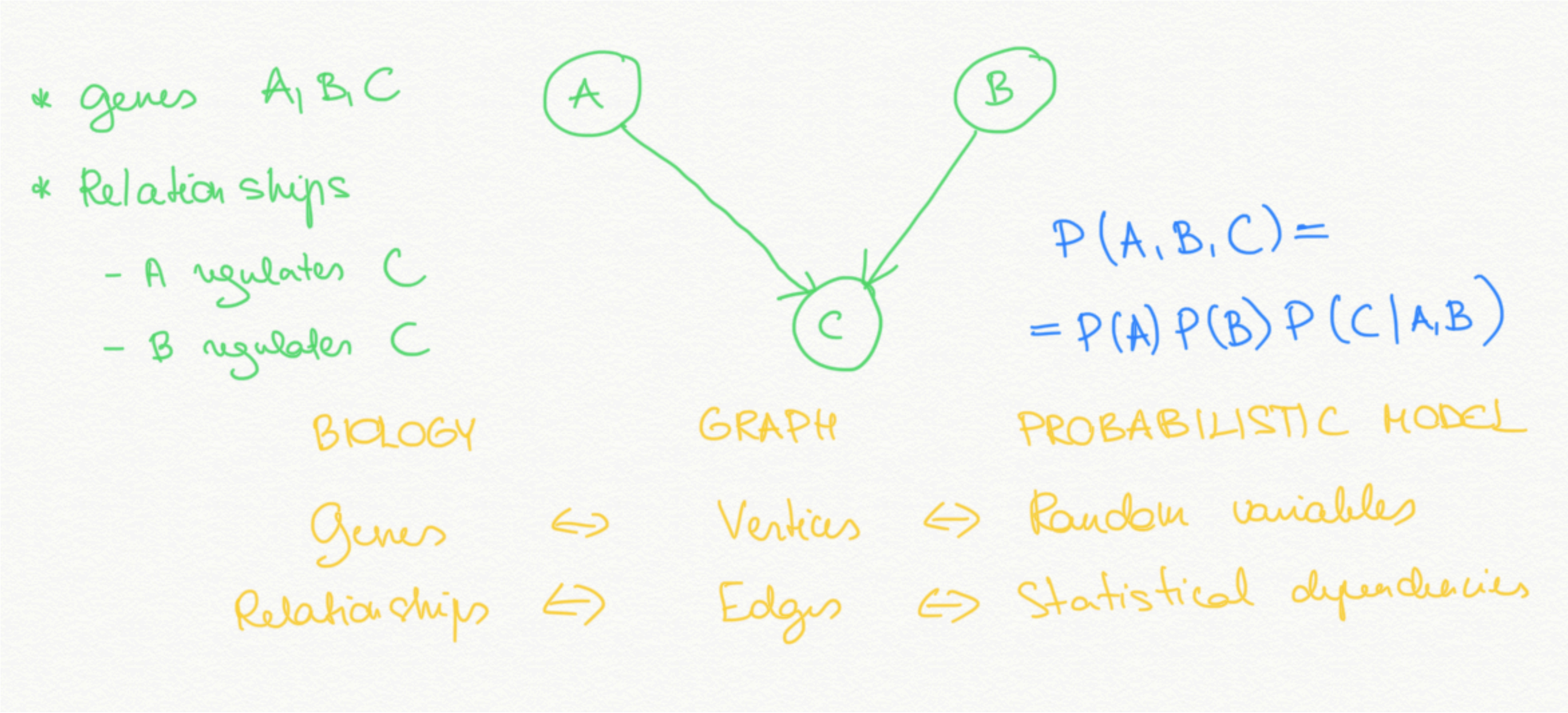
Undirected graphical models

Kaie Kubjas, 25.11.2020

Agenda

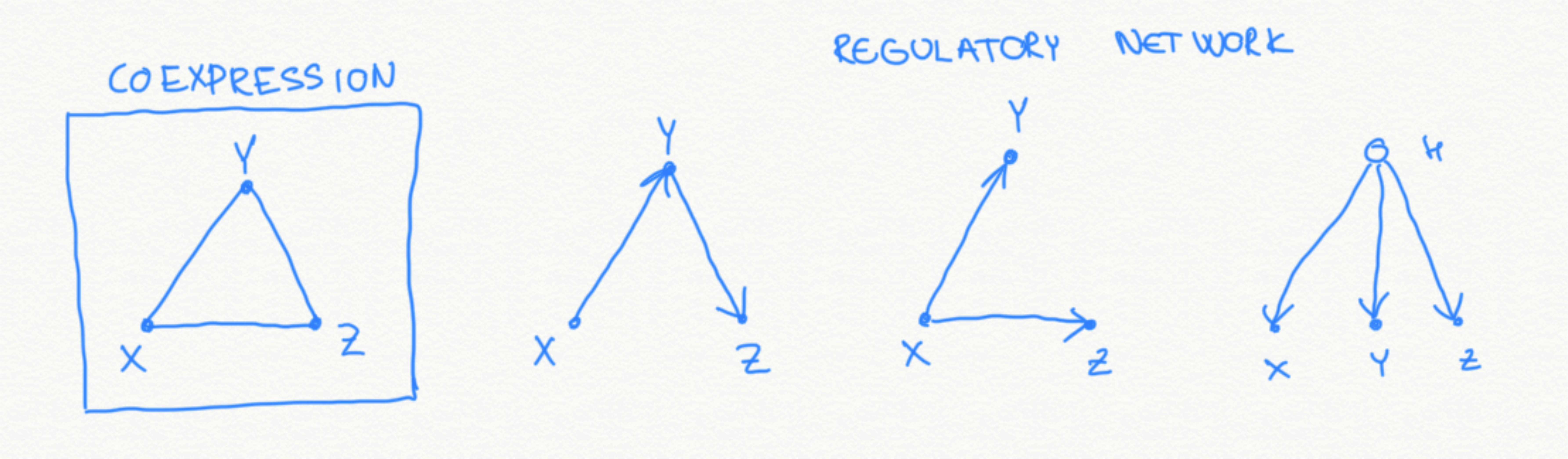
- How a graph encodes conditional independence statements
- When a conditional independence ideal is equal to a parametrized graphical model
- This lecture will connect
  - monomial parametrizations of discrete exponential families
  - toric ideals
  - conditional independence (ideals)
- Next time: Maximum likelihood estimation for undirected graphical models

Graphical models example



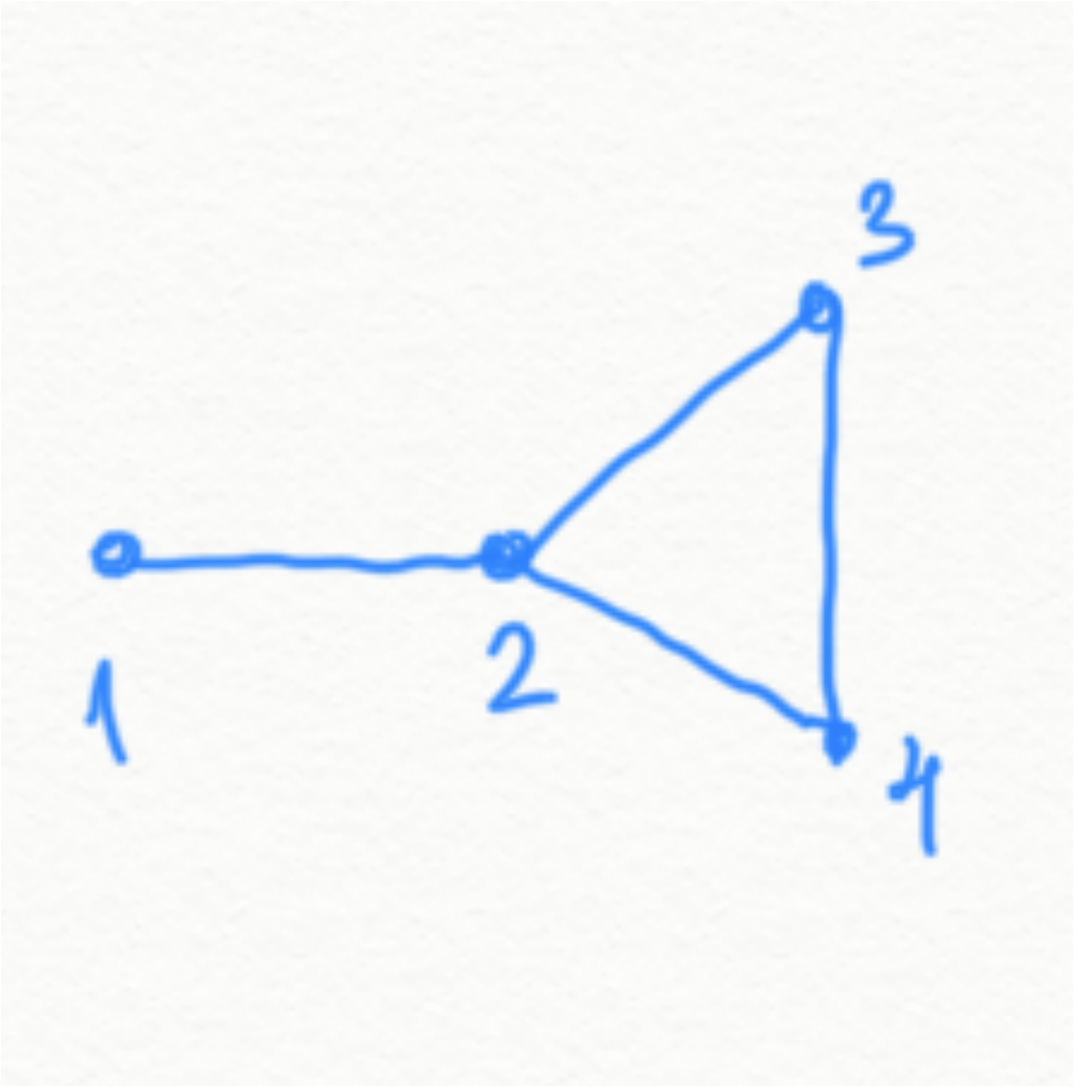
Correlation vs causation

- Genes regulated as $X \rightarrow Y \rightarrow Z$
- X and Z are correlated, but do not interact directly



Graphs

- Graph $G = (V, E)$
    - Nodes or vertices V
    - Edges $E \subseteq V \times V$
- A graph is undirected if $(u, v) \in E$ implies that $(v, u) \in E$
- Corresponding random vector $X = (X_v : v \in V)$



Graphical models

In the graphical model associated to a graph G, an edge (u, v) of the graph G expresses some sort of dependence between the vertices u and v.

Separator

- A path between vertices u and w in a graph G is a sequence of vertices $u = v_1, v_2, \ldots, v_k = w$ such that each $(v_{i-1}, v_i) \in E$.

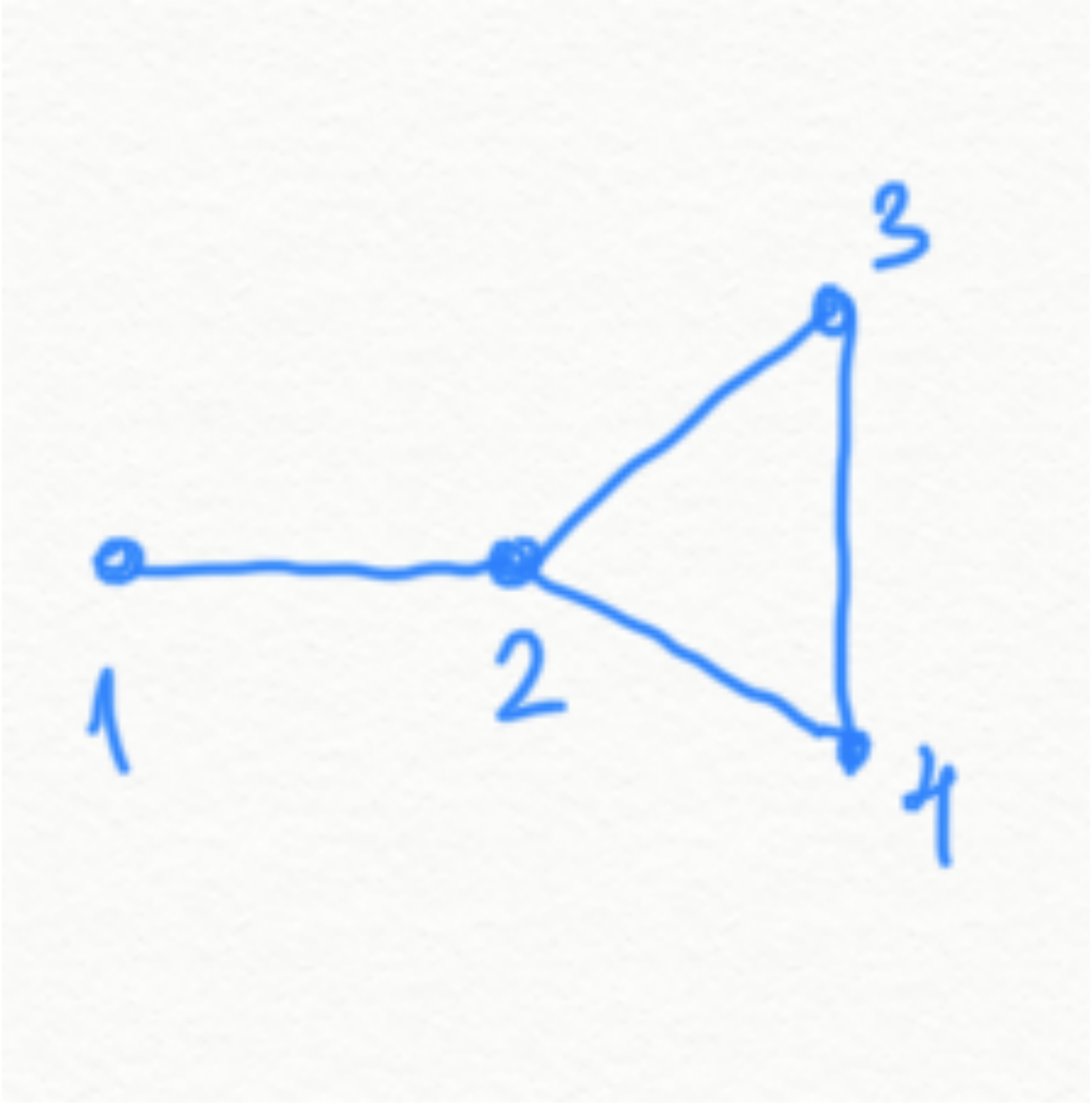
- A pair of vertices $a, b \in V$ is separated by a set of vertices $C \subseteq V \backslash \{a, b\}$ if every path from a to b contains a vertex in C.

- Let A, B, C be disjoint subsets of V. Then A and B are separated by C, if a and b are separated by C for any $a \in A$ and $b \in B$.

Separator

Poll: Let G be a graph with nodes $\{1,2,3,4\}$ and edges $(1,2), (2,3), (2,4), (3,4)$. Which of the following sets are separators for the nodes 1 and 4?

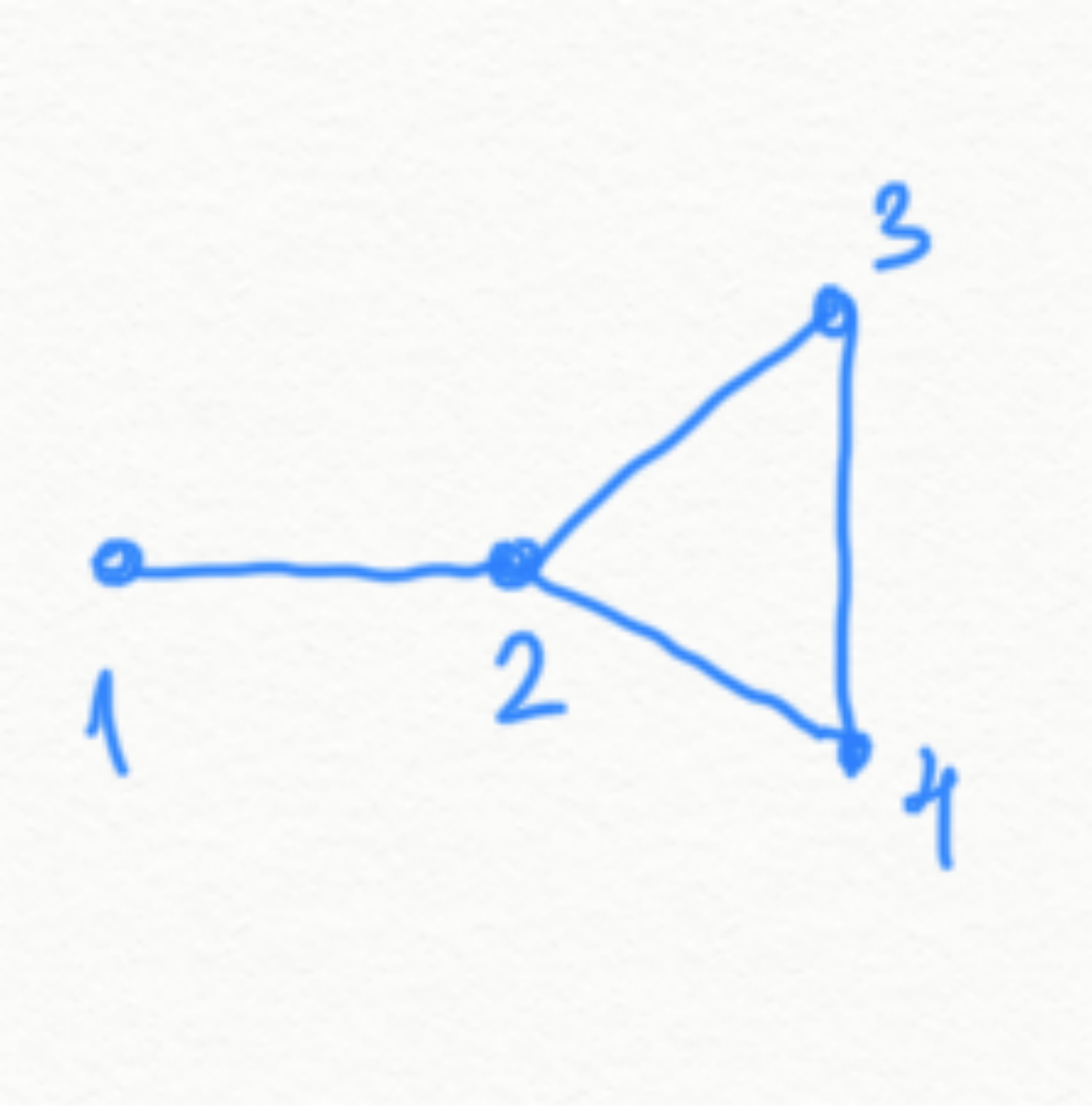
1. $\{2\}$
2. $\{3\}$
3. $\{2,3\}$
4. $\{1,2,3,4\}$



Separator

Poll: Let G be a graph with nodes $\{1,2,3,4\}$ and edges $(1,2), (2,3), (2,4), (3,4)$. Which of the following sets are separators for the nodes 1 and 4?

1. $\{2\}$ - Correct
2. $\{3\}$
3. $\{2,3\}$ - Correct
4. $\{1,2,3,4\}$



Conditional independence

Def: Let $A, B, C \subseteq [m]$ be pairwise disjoint subsets. We say that X_A is conditionally independent of X_B given X_C if and only if

$$f_{A\cup B|C}(x_A, x_B \,|\, x_C) = f_{A|C}(x_A \,|\, x_C) f_{B|C}(x_B \,|\, x_C)$$

for all x_A, x_B, x_C.

- The notation $X_A \perp\!\!\!\perp X_B \,|\, X_C$ (or $A \perp\!\!\!\perp B \,|\, C$) denotes that the random vector X satisfies the conditional independence (CI) statement that X_A is conditionally independent of X_B given X_C .

Pairwise Markov property

Let $G = (V, E)$ be an undirected graph.

Def: The pairwise Markov property associated to G consists of all conditional independence statements $X_u \perp\!\!\!\perp X_v \mid X_{V \setminus \{u,v\}}$, where (u, v) is not an edge of G.

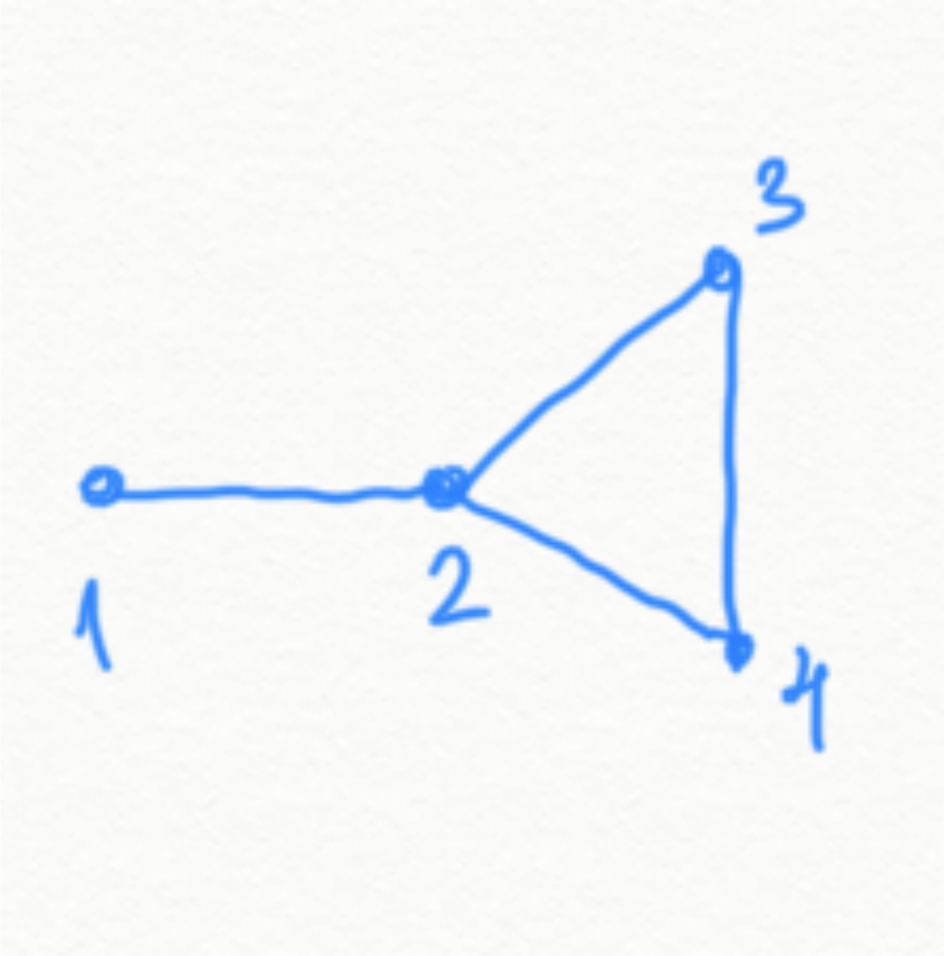
Example: The pairwise Markov property associated to G is:

1. $\{1 \perp\!\!\!\perp 3 \mid (2,4), 1 \perp\!\!\!\perp 4 \mid (2,3)\}$

2. $\{1 \perp\!\!\!\perp 3 \mid 2, 1 \perp\!\!\!\perp 4 \mid 2\}$

3. $\{1 \perp\!\!\!\perp 3 \mid (2,4)\}$

4. $\{1 \perp\!\!\!\perp 4 \mid (2,3)\}$



Pairwise Markov property

Let $G = (V, E)$ be an undirected graph.

<u>Def:</u> The pairwise Markov property associated to G consists of all conditional independence statements $X_u \perp\!\!\!\perp X_v \mid X_{V \setminus \{u,v\}}$, where (u, v) is not an edge of G.

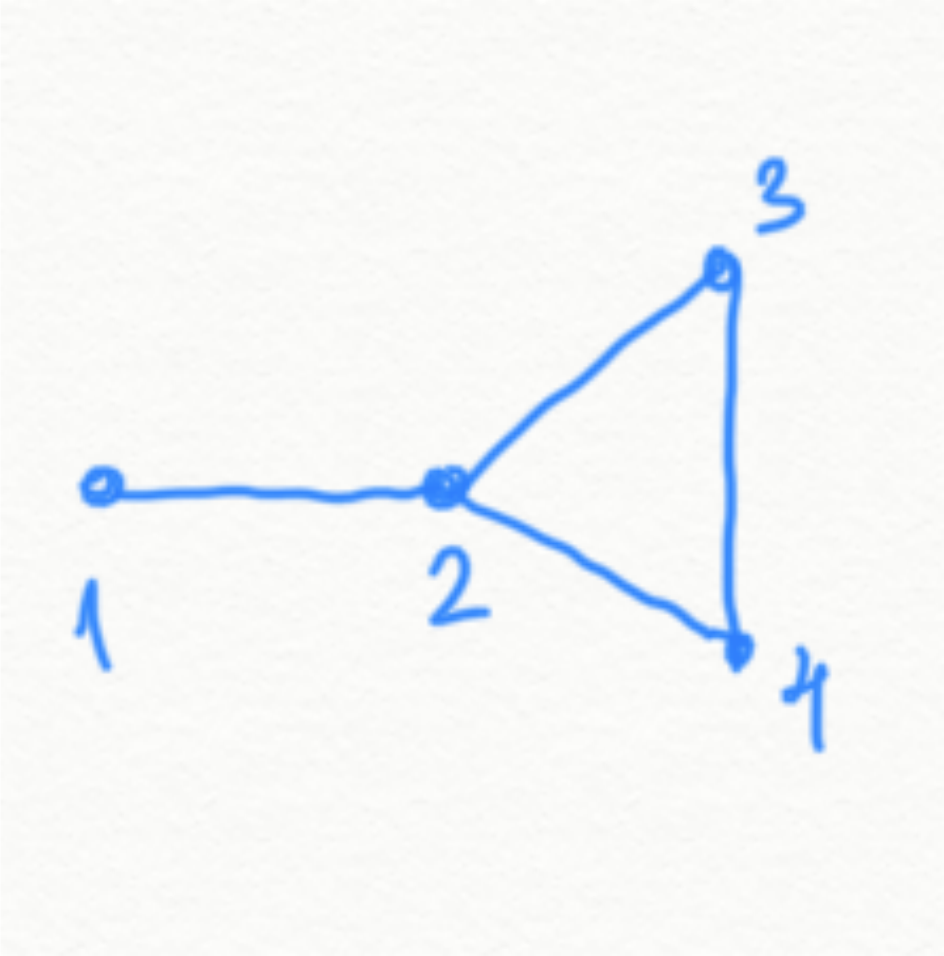
<u>Example:</u> The pairwise Markov property associated to G is:

1. $\{1 \perp\!\!\!\perp 3 \mid (2,4), 1 \perp\!\!\!\perp 4 \mid (2,3)\}$ - Correct

2. $\{1 \perp\!\!\!\perp 3 \mid 2, 1 \perp\!\!\!\perp 4 \mid 2\}$

3. $\{1 \perp\!\!\!\perp 3 \mid (2,4)\}$

4. $\{1 \perp\!\!\!\perp 4 \mid (2,3)\}$



Multivariate Gaussian random variables

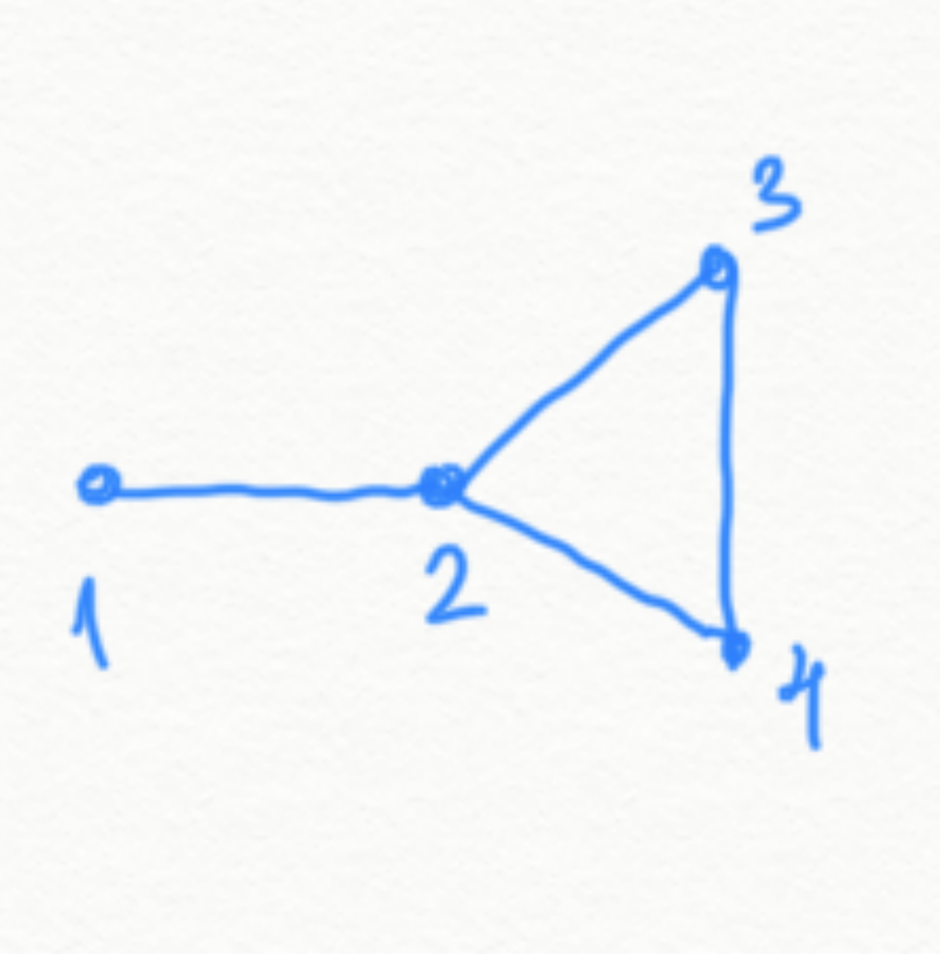
- The CI statement $X_u \perp\!\!\!\perp X_v \mid X_{V \setminus \{u,v\}}$ is equivalent to the matrix $\Sigma_{V \setminus \{u\}, V \setminus \{v\}}$ having rank $|V \setminus \{u, v\}|$ or equivalently $\det(\Sigma_{V \setminus \{u\}, V \setminus \{v\}}) = 0$.
- This is equivalent to $(\Sigma^{-1})_{u,v} = 0$.
- The pairwise Markov property holds for a Gaussian distribution if and only if the entries of the concentration matrix corresponding to non-edges are zero.

Multivariate Gaussian random variables

Which form do the concentration matrices of a Gaussian distribution obeying the pairwise Markov property have?

1. $$\begin{pmatrix} k_{11} & 0 & k_{13} & k_{14} \\ 0 & k_{22} & 0 & 0 \\ k_{13} & 0 & k_{33} & 0 \\ k_{14} & 0 & 0 & k_{44} \end{pmatrix}$$

2. $$\begin{pmatrix} k_{11} & k_{12} & 0 & 0 \\ k_{12} & k_{22} & k_{23} & k_{24} \\ 0 & k_{23} & k_{33} & k_{34} \\ 0 & k_{24} & k_{34} & k_{44} \end{pmatrix}$$

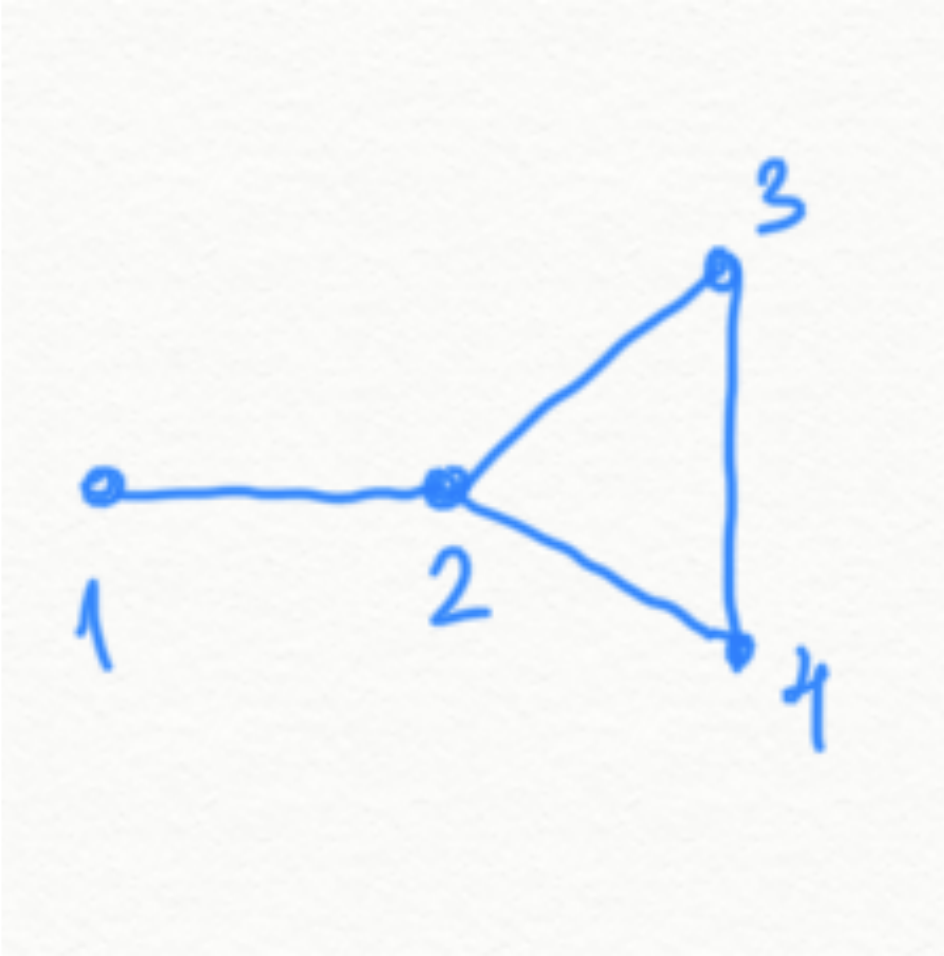


Multivariate Gaussian random variables

Which form do the concentration matrices of a Gaussian distribution obeying the pairwise Markov property have?

1. $$\begin{pmatrix} k_{11} & 0 & k_{13} & k_{14} \\ 0 & k_{22} & 0 & 0 \\ k_{13} & 0 & k_{33} & 0 \\ k_{14} & 0 & 0 & k_{44} \end{pmatrix}$$

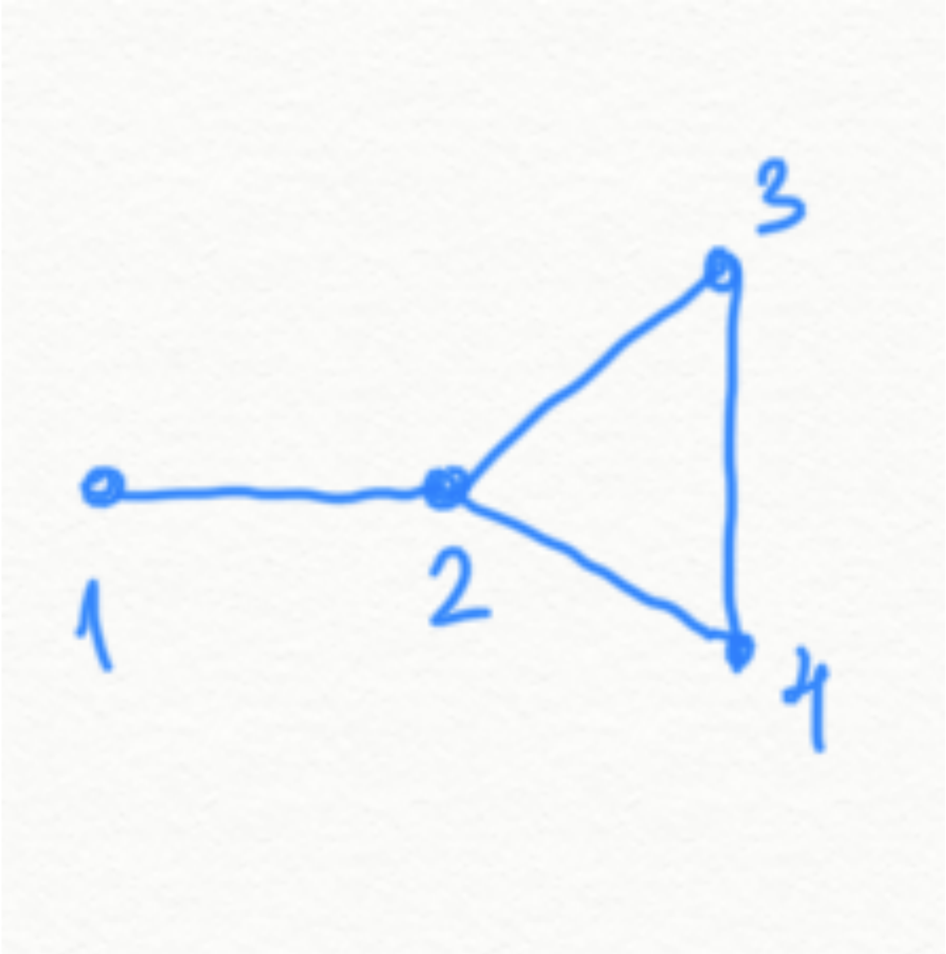
2. $$\begin{pmatrix} k_{11} & k_{12} & 0 & 0 \\ k_{12} & k_{22} & k_{23} & k_{24} \\ 0 & k_{23} & k_{33} & k_{34} \\ 0 & k_{24} & k_{34} & k_{44} \end{pmatrix}$$ - Correct



Global Markov property

Def: The global Markov property associated to G consists of all conditional independence statements $X_A \perp\!\!\!\perp X_B \mid X_C$ for all disjoint sets A, B, and C such that C separates A and B in G.

Example: The global Markov property associated to G is:



1. $\{1 \perp\!\!\!\perp (3,4) \mid 2\}$

2. $\{1 \perp\!\!\!\perp 3 \mid (2,4), 1 \perp\!\!\!\perp 4 \mid (2,3)\}$

3. $\{1 \perp\!\!\!\perp 3 \mid (2,4), 1 \perp\!\!\!\perp 4 \mid (2,3), 1 \perp\!\!\!\perp (3,4) \mid 2\}$

Global Markov property

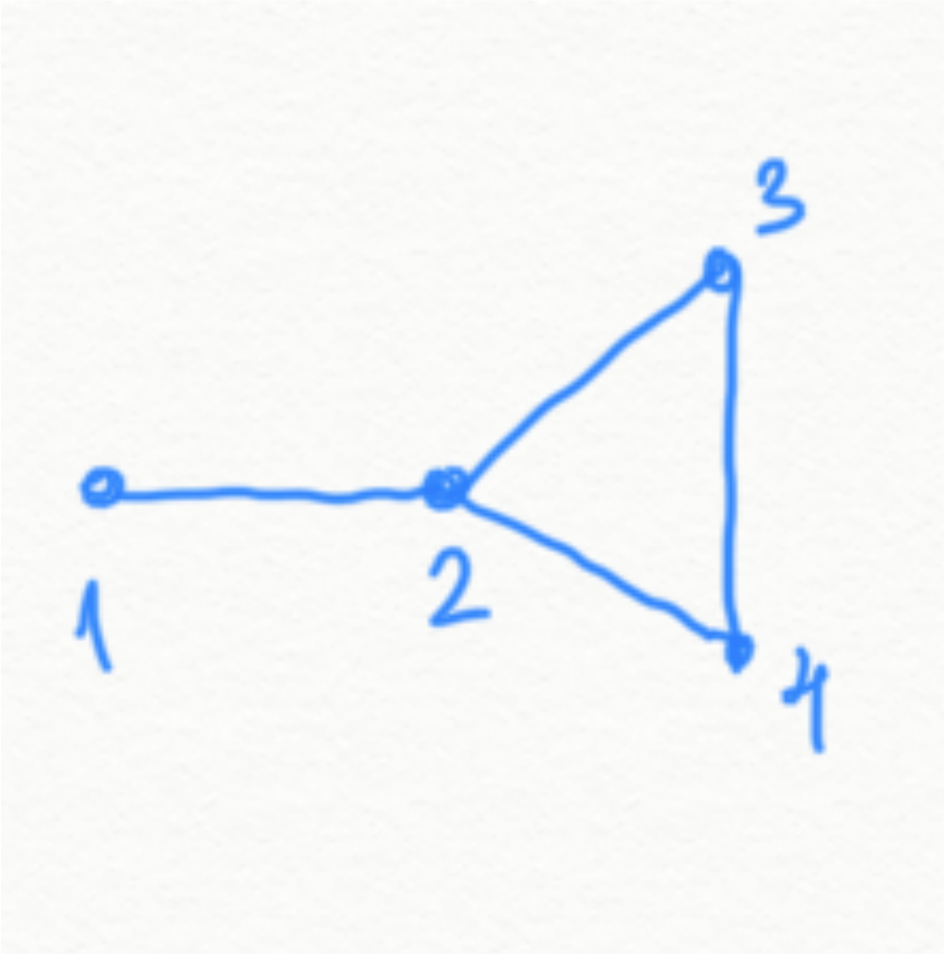
Def: The global Markov property associated to G consists of all conditional independence statements $X_A \perp\!\!\!\perp X_B \mid X_C$ for all disjoint sets A, B, and C such that C separates A and B in G.

Example: The global Markov property associated to G is:

1. $\{1 \perp\!\!\!\perp (3,4) \mid 2\}$

2. $\{1 \perp\!\!\!\perp 3 \mid (2,4), 1 \perp\!\!\!\perp 4 \mid (2,3)\}$

3. $\{1 \perp\!\!\!\perp 3 \mid (2,4), 1 \perp\!\!\!\perp 4 \mid (2,3), 1 \perp\!\!\!\perp (3,4) \mid 2\}$ - Correct



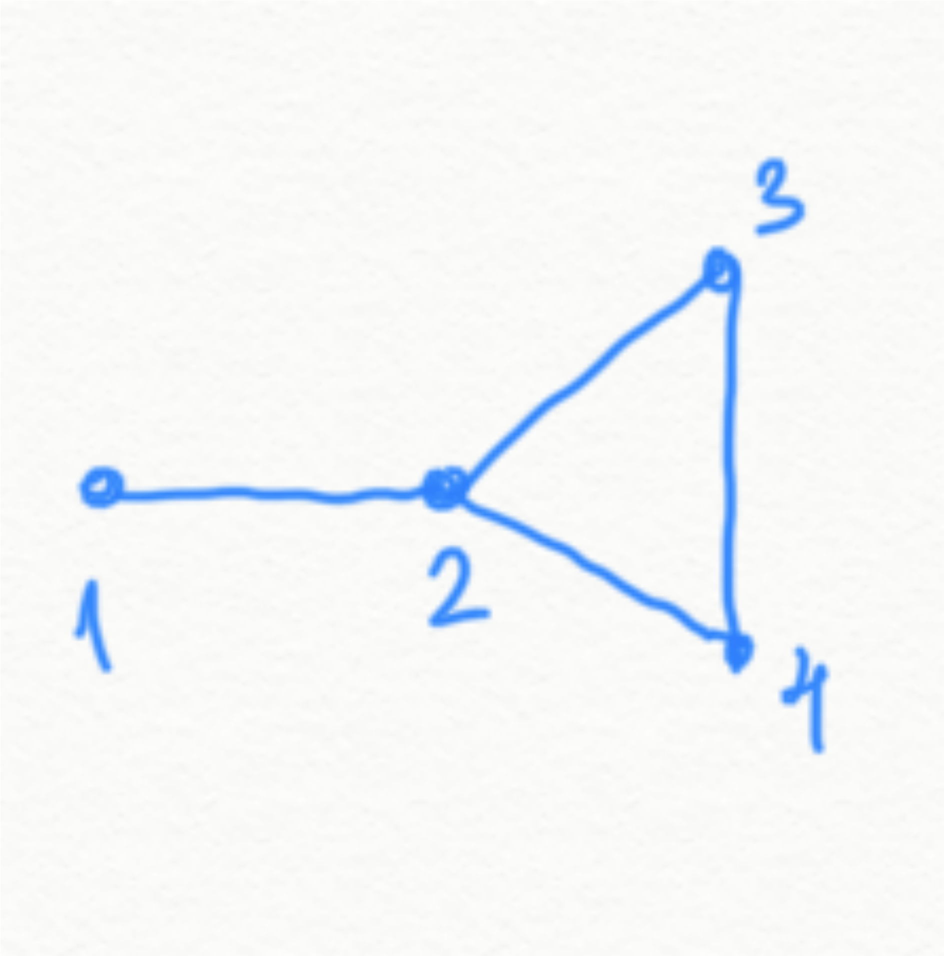
Markov properties

- It always holds $\mathscr{C}_{pairs} \subseteq \mathscr{C}_{global}$.

Example:

- $\mathscr{C}_{pairs} = \{1 \perp\!\!\!\perp 3 \,|\, (2,4), 1 \perp\!\!\!\perp 4 \,|\, (2,3)\}$

- $\mathscr{C}_{\text{global}} = \mathscr{C}_{\text{pairs}} \cup \{1 \perp\!\!\!\perp (3,4) \,|\, 2\}$



Intersection axiom

Prop (Intersection axiom): Suppose that $f(x) > 0$ for all x. Then

$$X_A \perp\!\!\!\perp X_B \,|\, X_{C \cup D} \text{ and } X_A \perp\!\!\!\perp X_C \,|\, X_{B \cup D} \implies X_A \perp\!\!\!\perp X_{B \cup C} \,|\, X_D.$$

- The condition $f(x) > 0$ for all x is stronger than necessary.
- For discrete random variables, precise conditions can be given which guarantee that the intersection axiom holds. This is done using algebra!

Markov properties

Theorem: If the distribution P of a random vector X satisfies the intersection axiom, then P obeys the pairwise Markov property for G if and only if it obeys the global Markov property for G.

Multivariate Gaussian random variables

For multivariate Gaussian random variables with non-singular covariance matrix, the density function is strictly positive.

$\Longrightarrow$ the intersection axiom holds

$\Longrightarrow$ the Markov properties are equivalent in this class of distributions

Factorization property

- Next we want to characterize all the distributions that satisfy the Markov properties for a given graph.

- Hammersley-Clifford theorem relates the implicit description of a graphical model through Markov properties to a parametric description.

Factorization property

- Let $G = (V, E)$ be an undirected graph.
- A subset of vertices $C \subseteq V$ is a clique if $(i, j) \in E$ for all $i, j \in C$.
- The set of maximal cliques of G is denoted $\mathscr{C}(G)$.
- For each $C \in \mathscr{C}(G)$, we introduce a continuous nonnegative potential function $\phi_C : \mathscr{X}_C \to \mathbb{R}_{\geq 0}$.

Maximal cliques

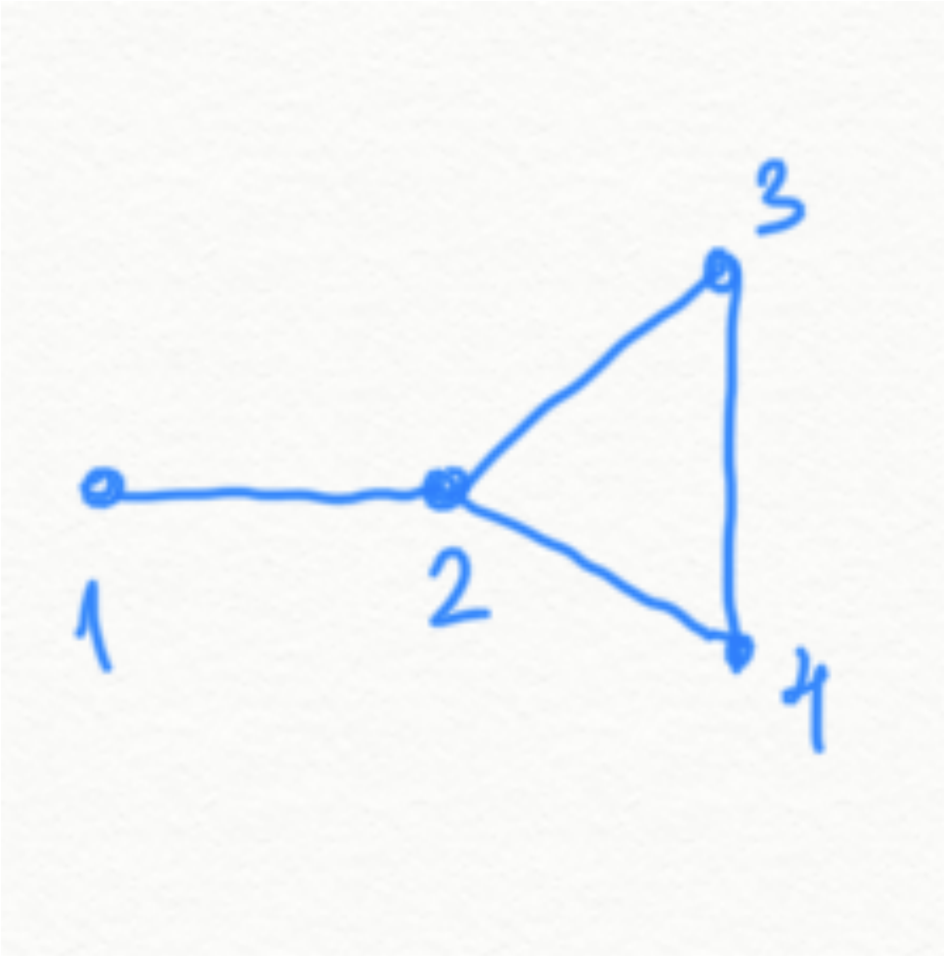
<u>Example:</u> Which are maximal cliques of G?

1. $\{1\}$

2. $\{1,2\}$

3. $\{1,2,3\}$

4. $\{2,3,4\}$



Maximal cliques

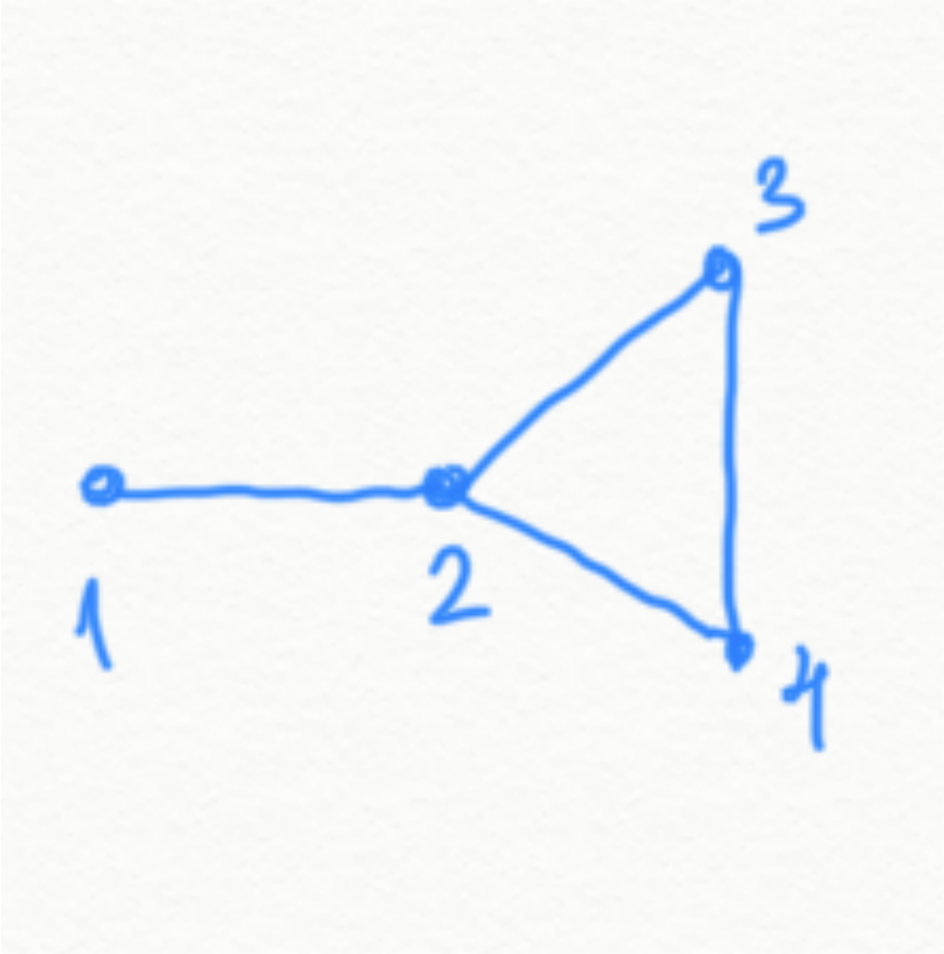
<u>Example:</u> Which are maximal cliques of G?

1. $\{1\}$

2. $\{1,2\}$ - Correct

3. $\{1,2,3\}$

4. $\{2,3,4\}$ - Correct



Factorization property

Def: The distribution of X factorizes according to the graph G if its probability density function $f(x)$ can be written as

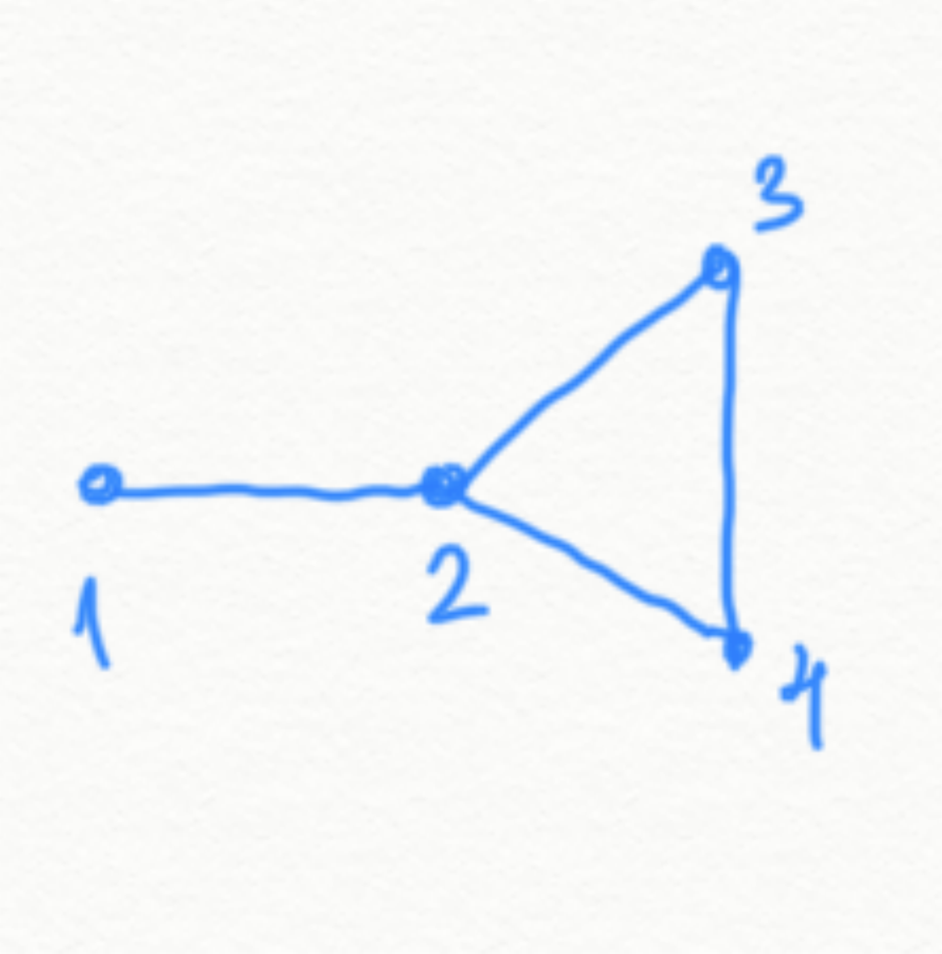
$$f(x) = \frac{1}{Z} \prod_{C \in \mathscr{C}(G)} \phi_C(x_C),$$

where ϕ_C are some potential functions and $Z < \infty$ is the normalizing constant.

Factorization property

$$f(x) = \frac{1}{Z} \prod_{C \in \mathscr{C}(G)} \phi_C(x_C)$$

Example: A distribution factorizes according to G if its density $f(x)$ can be written as



$$f(x) = \frac{1}{Z}\phi_{12}(x_1, x_2)\phi_{234}(x_2, x_3, x_4).$$

Hammersley-Clifford

Theorem (Hammersley-Clifford): A distribution with positive and continuous density f satisfies the pairwise Markov property on the graph G if and only if it factorizes according to G.

- The Gaussian case is completely covered by the Hammersley-Clifford theorem.
- All distributions on a discrete space are considered continuous.
- What happens in the discrete case?

Discrete distributions

- Let X be a discrete random vector with state space $\mathscr{R} = \prod_{j=1}^{m} [r_j]$.
- Write $i_C := (i_j)_{j \in C} \in R_C$.
- Then we can write $\phi_C(x_C)$ as $\theta^{(C)}_{i_C}$.
- $f(x) = \frac{1}{Z}\phi_{12}(x_1, x_2)\phi_{234}(x_2, x_3, x_4)$ becomes $p_{i_1 i_2 i_3 i_4} = \frac{1}{Z}\theta^{(12)}_{i_1 i_2}\theta^{(234)}_{i_2 i_3 i_4}$

Discrete distributions

- The distribution p of X factors according to G if

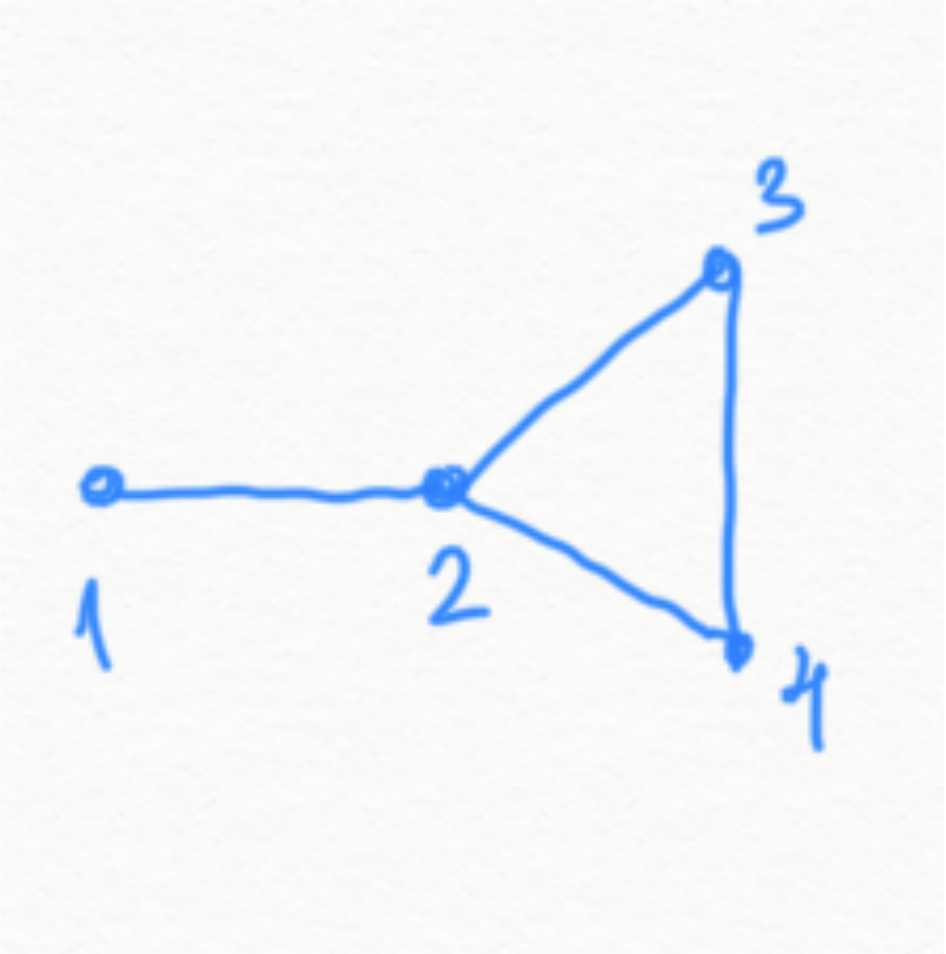
$$p_{i_1 i_2 \cdots i_m} = \frac{1}{Z(\theta)} \prod_{C \in \mathscr{C}(G)} \theta^{(C)}_{i_C},$$

which is a monomial parametrization.

- Hence the set of distributions that factorize according to a graph G form a hierarchical log-linear model.

- We will denote this model by I_G.

Discrete distributions



- $\mathscr{C}_{\text{pairs}} = \{1 \perp\!\!\!\perp 3 \,|\, (2,4), 1 \perp\!\!\!\perp 4 \,|\, (2,3)\}$

- $\mathscr{C}_{\text{global}} = \mathscr{C}_{\text{pairs}} \cup \{1 \perp\!\!\!\perp (3,4) \,|\, 2\}$

- $p(x) = \dfrac{1}{Z}\theta^{(12)}_{i_1 i_2}\theta^{(234)}_{i_2 i_3 i_4}$

Discrete conditional independence models

Prop: If X is a discrete random vector, then the conditional independence statement $X_A \perp\!\!\!\perp X_B \mid X_C$ holds if and only if

$$p_{i_A,i_B,i_C,+} \cdot p_{j_A,j_B,i_C,+} - p_{i_A,j_B,i_C,+} \cdot p_{j_A,i_B,i_C,+} = 0$$

for all $i_A, j_A \in \mathcal{R}_A$, $i_B, j_B \in \mathcal{R}_B$ and $i_C \in \mathcal{R}_C$.

- The notation $p_{i_A,i_B,i_C,+}$ denotes the probability $P(X_A = i_A, X_B = i_B, X_C = i_C)$ which can be written as

$$p_{i_A,i_B,i_C,+} = \sum_{j_{[m]\setminus A\cup B\cup C} \in \mathcal{R}_{[m]\setminus A\cup B\cup C}} p_{i_A,i_B,i_C,j_{[m]\setminus A\cup B\cup C}}.$$

Pairwise Markov property

- $\mathscr{C}_{\text{pairs}} = \{1 \perp\!\!\!\perp 3 \,|\, (2,4), 1 \perp\!\!\!\perp 4 \,|\, (2,3)\}$
- Poll: How many polynomials generate the corresponding CI ideal?
- $M_1 = \begin{pmatrix} p_{0000} & p_{0001} & p_{0010} & p_{0011} \\ p_{1000} & p_{1001} & p_{1010} & p_{1011} \end{pmatrix}$
- $M_2 = \begin{pmatrix} p_{0100} & p_{0101} & p_{0110} & p_{0111} \\ p_{1100} & p_{1101} & p_{1110} & p_{1111} \end{pmatrix}$
- The conditional independence ideal for each statement is generated by two minors of M_1 and two minors of M_2

```
i1 : R1 = QQ[p_(0,0,0,0)..p_(1,1,1,1)]

o1 = R1

o1 : PolynomialRing

i2 : M1 = matrix{{p_(0,0,0,0),p_(0,0,0,1),p_(0,0,1,0),p_(0,0,1,1)},{p_(1,0,0,0),p_(1,0,0,1),p_(1,0,1,0),p_(1,0,1,1)}}

o2 = | p_(0,0,0,0) p_(0,0,0,1) p_(0,0,1,0) p_(0,0,1,1) |
     | p_(1,0,0,0) p_(1,0,0,1) p_(1,0,1,0) p_(1,0,1,1) |

               2        4
o2 : Matrix R1  <--- R1

i3 : M2 = matrix{{p_(0,1,0,0),p_(0,1,0,1),p_(0,1,1,0),p_(0,1,1,1)},{p_(1,1,0,0),p_(1,1,0,1),p_(1,1,1,0),p_(1,1,1,1)}}

o3 = | p_(0,1,0,0) p_(0,1,0,1) p_(0,1,1,0) p_(0,1,1,1) |
     | p_(1,1,0,0) p_(1,1,0,1) p_(1,1,1,0) p_(1,1,1,1) |

               2        4
o3 : Matrix R1  <--- R1

i4 : IP = ideal(det(M1_{0,2}),det(M1_{1,3}),det(M2_{0,2}),det(M2_{1,3}),det(M1_{0,1}),det(M1_{2,3}),det(M2_{0,1}),det(M2_{2,3}))

o4 = ideal (- p        p         + p        p        , - p        p         +
               0,0,1,0 1,0,0,0    0,0,0,0 1,0,1,0       0,0,1,1 1,0,0,1
     ------------------------------------------------------------------------
     p        p        , - p        p         + p        p        , -
      0,0,0,1 1,0,1,1       0,1,1,0 1,1,0,0    0,1,0,0 1,1,1,0
     ------------------------------------------------------------------------
     p        p         + p        p        , - p        p         +
      0,1,1,1 1,1,0,1      0,1,0,1 1,1,1,1       0,0,0,1 1,0,0,0
     ------------------------------------------------------------------------
     p        p        , - p        p         + p        p        , -
      0,0,0,0 1,0,0,1       0,0,1,1 1,0,1,0    0,0,1,0 1,0,1,1
     ------------------------------------------------------------------------
     p        p         + p        p        , - p        p         +
      0,1,0,1 1,1,0,0      0,1,0,0 1,1,0,1       0,1,1,1 1,1,1,0
     ------------------------------------------------------------------------
     p        p        )
      0,1,1,0 1,1,1,1

o4 : Ideal of R1
```

Global Markov property

- $\mathscr{C}_{\text{global}} = \mathscr{C}_{\text{pairs}} \cup \{1 \perp\!\!\!\perp (3,4) \,|\, 2\}$
- $M_1 = \begin{pmatrix} p_{0000} & p_{0001} & p_{0010} & p_{0011} \\ p_{1000} & p_{1001} & p_{1010} & p_{1011} \end{pmatrix}$
- $M_2 = \begin{pmatrix} p_{0100} & p_{0101} & p_{0110} & p_{0111} \\ p_{1100} & p_{1101} & p_{1110} & p_{1111} \end{pmatrix}$
- The conditional independence ideal $\mathscr{C}_{\text{global}}$ is generated by all 2×2 minors of M_1 and M_2

Factorization according to G

- $p_{i_1 i_2 i_3 i_4} = \frac{1}{Z} \theta^{(12)}_{i_1 i_2} \theta^{(234)}_{i_2 i_3 i_4}$
- Poll: How many parameters does this parametrization map have?
- $p_{ijkl} = a_{ij} b_{jkl}$
- We obtain the toric ideal I_G by eliminating the variables a_{ij}, b_{jkl}:

$$I_G = \langle p_{ijkl} - a_{ij} b_{jkl} : (i, j, k, l) \in \{0,1\}^4 \rangle \cap \mathbb{R}[p]$$

```
i6 : R3 = QQ[p_(0,0,0,0)..p_(1,1,1,1),a_(0,0)..a_(1,1),b_(0,0,0)..b_(1,1,1)]

o6 = R3

o6 : PolynomialRing

i7 : IF = ideal flatten flatten flatten for i to 1 list for j to 1 list for k to 1 list for l to 1 list p_(i,j,k,l)-a_(i,j)*b_(j,k,l)

o7 = ideal (- a   b      + p        , - a   b      + p        , - a   b      +
              0,0 0,0,0     0,0,0,0      0,0 0,0,1     0,0,0,1      0,0 0,1,0
     ------------------------------------------------------------------------
     p        , - a   b      + p        , - a   b      + p        , - a   b      +
      0,0,1,0      0,0 0,1,1     0,0,1,1      0,1 1,0,0     0,1,0,0      0,1 1,0,1
     ------------------------------------------------------------------------
     p        , - a   b      + p        , - a   b      + p        , - a   b      +
      0,1,0,1      0,1 1,1,0     0,1,1,0      0,1 1,1,1     0,1,1,1      1,0 0,0,0
     ------------------------------------------------------------------------
     p        , - a   b      + p        , - a   b      + p        , - a   b      +
      1,0,0,0      1,0 0,0,1     1,0,0,1      1,0 0,1,0     1,0,1,0      1,0 0,1,1
     ------------------------------------------------------------------------
     p        , - a   b      + p        , - a   b      + p        , - a   b      +
      1,0,1,1      1,1 1,0,0     1,1,0,0      1,1 1,0,1     1,1,0,1      1,1 1,1,0
     ------------------------------------------------------------------------
     p        , - a   b      + p        )
      1,1,1,0      1,1 1,1,1     1,1,1,1

o7 : Ideal of R3

i8 : JF = eliminate(IF,join(toList(a_(0,0)..a_(1,1)), toList(b_(0,0,0)..b_(1,1,1))))

o8 = ideal (p        p        - p        p        , p        p        -
             0,1,1,1 1,1,1,0    0,1,1,0 1,1,1,1    0,1,1,1 1,1,0,1
     ------------------------------------------------------------------------
     p        p        , p        p        - p        p        , p        p        -
      0,1,0,1 1,1,1,1    0,1,1,0 1,1,0,1    0,1,0,1 1,1,1,0    0,1,1,1 1,1,0,0
     ------------------------------------------------------------------------
     p        p        , p        p        - p        p        , p        p        -
      0,1,0,0 1,1,1,1    0,1,1,0 1,1,0,0    0,1,0,0 1,1,1,0    0,1,0,1 1,1,0,0
     ------------------------------------------------------------------------
     p        p        , p        p        - p        p        , p        p        -
      0,1,0,0 1,1,0,1    0,0,1,1 1,0,1,0    0,0,1,0 1,0,1,1    0,0,1,1 1,0,0,1
     ------------------------------------------------------------------------
     p        p        , p        p        - p        p        , p        p        -
      0,0,0,1 1,0,1,1    0,0,1,0 1,0,0,1    0,0,0,1 1,0,1,0    0,0,1,1 1,0,0,0
     ------------------------------------------------------------------------
     p        p        , p        p        - p        p        , p        p        -
      0,0,0,0 1,0,1,1    0,0,1,0 1,0,0,0    0,0,0,0 1,0,1,0    0,0,0,1 1,0,0,0
     ------------------------------------------------------------------------
     p        p        )
      0,0,0,0 1,0,0,1

o8 : Ideal of R3
```

Comparison of ideals

In this example:

- $I_G = I_{\text{global}(G)}$
- $I_{\text{pairwise}(G)}$ is different
- $I_{\text{pairwise}(G)}$ has 9 primary components, one of them is $I_G = I_{\text{global}(G)}$
- Each of the other eight components contains at least one variable p_{ijkl}
- This means that the corresponding irreducible varieties intersect the boundary of the probability simplex Δ_{15}

Comparison of ideals

- This shows that the positivity assumption in the Hammersley-Clifford Theorem is necessary

- One primary component is $\langle p_{0,0,0,0}, p_{1,0,0,0}, p_{1,0,1,1}, p_{0,0,1,1}, p_{1,1,0,0}, p_{0,1,0,0}, p_{0,1,1,1}, p_{1,1,1,1} \rangle$

- It represents the family of distributions such that $P(X_3 = X_4) = 1$.

- All such distributions satisfy the pairwise Markov property, but they are not in the model characterized by G.

Comparison of ideals

- In the previous example, the polynomials implied by the global Markov property characterize I_G.

- This is not true in general.

- A graph G is chordal if every induced cycle of length 4 or larger has a chord.

Theorem: $I_G = I_{\text{global}(G)}$ if and only if G is a chordal graph.

Conclusion

- Implicit description of an undirected graphical model through Markov properties
- Parametric description of an undirected graphical model through factorization according to a graph
- Hammersley-Clifford theorem when a graphical model is given by pairwise Markov properties
- The failure of the Hammersley-Clifford theorem

Next time

- Maximum likelihood estimation for undirected graphical models
- Bachelor and Master thesis topics presentation

Literature

- Lauritzen “Graphical Models”
- Maathuis, Drton, Lauritzen, Wainwright “Handbook of Graphical Models”
- Koller and Friedman “Probabilistic Graphical Models”
- Peters, Danzig, Schölkopf “Elements of Causal Inference”