Graphical model formalism, factorization properties and conditional independance.

Guillaume Obozinski

Ecole des Ponts - ParisTech

Master MVA, course on Probabilistic Graphical Models

Independence concepts

Independence: $X \perp\!\!\!\perp Y$

We say that X et Y are independents and write $X \perp\!\!\!\perp Y$ ssi:

$$\forall x, y, \qquad P(X = x, Y = y) = P(X = x)\, P(Y = y)$$

Conditional Independence: $X \perp\!\!\!\perp Y \mid Z$

- On says that X and Y are independent conditionally on Z and
- write $X \perp\!\!\!\perp Y \mid Z$ iff:

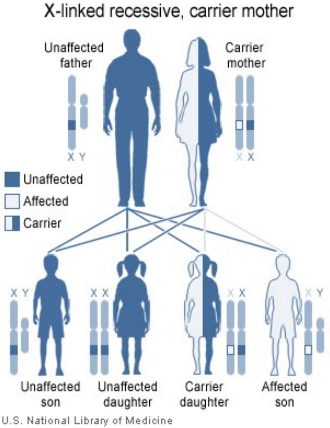
$\forall x, y, z,$

$$P(X = x, Y = y \mid Z = z) = P(X = x | Z = z)\, P(Y = y | Z = z)$$

Conditional Independence exemple

Example of
"X-linked recessive inheritance":

Transmission of the gene
responsible for hemophilia

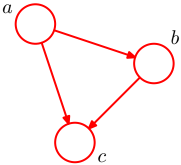


Risk for sons from an unaffected father:

- dependance between the situation of the two brothers.
- conditionally independent given that the mother is a carrier of the gene or not.

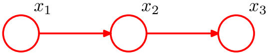
Directed graphical model or Bayesian network

$$p(a, b, c) = p(a)\, p(b|a)\, p(c|b, a)$$

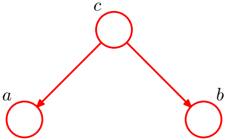


$$p(x_1, x_2) = p(x_1)p(x_2)$$

$$p(x_1, x_2, x_3) = p(x_1)p(x_2|x_1)p(x_3|x_2)$$



$$a \perp\!\!\!\perp b \mid c$$

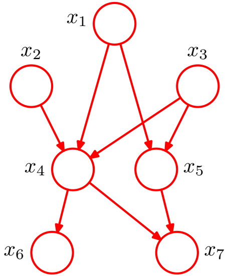


Directed graphical model or Bayesian network

Factorization according to a directed graph

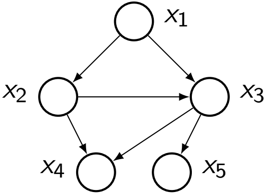
Definition: a distribution factorizes according to a directed graph

$$\prod_{j=1}^{p} p(x_j | x_{\Pi_j})$$



$$p(x_1) \prod_{j=2}^{M} p(x_j | x_{j-1})$$

How to parameterize an Oriented graphical model?



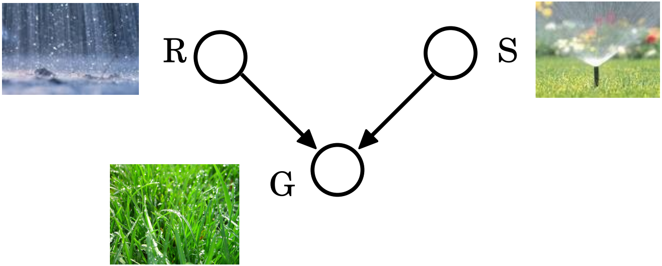
Conditional Probability tables

- $x_1 \in \{0, 1\}$
- $x_2 \in \{0, 1, 2\}$
- $x_3 \in \{0, 1, 2\}$

		$p(x_3 = k)$		
x_1	x_2	0	1	2
0	0	1	0	0
0	1	1	0	0
0	2	0.1	0	0.9
1	0	1	0	0
1	1	0.5	0.5	0
1	2	0.2	0.3	0.5

$$p(\mathbf{x}; \boldsymbol{\theta}) = p(x_1; \theta_1)\, p(x_2|x_1; \theta_2)\, p(x_3|x_2, x_1; \theta_3)\, p(x_4|x_3, x_2; \theta_4)\, p(x_5|x_3; \theta_5)$$

The Sprinkler



$$P(S=1)=0.5$$

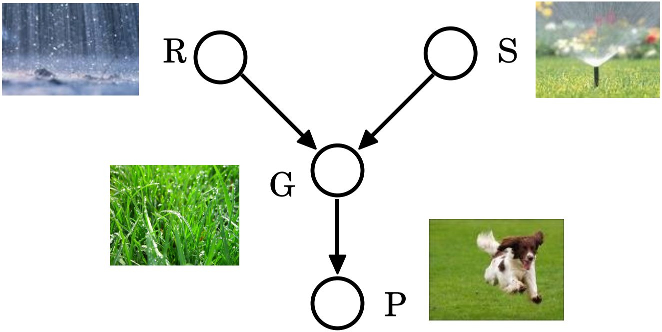
$$P(R=1)=0.2$$

- $R=1$: it has rained
- $S=1$: the sprinkler worked
- $G=1$: the grass is wet

$P(G=1\|S,R)$	R=0	R=1
S=0	0.01	0.8
S=1	0.8	0.95

- Given that we observe that the grass is wet, are R and S independent?

The Sprinkler II



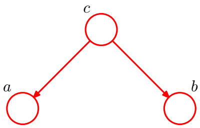
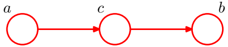
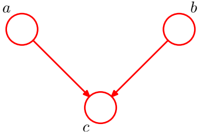
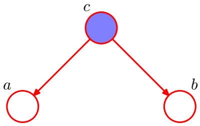
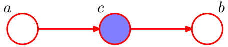
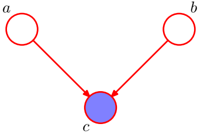
- $R = 1$: it has rained
- $S = 1$: the sprinkler worked
- $G = 1$: the grass is wet
- $P = 2$: the paws of the dog are wet

$P(S = 1) = 0.5 \quad P(R = 1) = 0.2$

$P(G = 1 \| S, R)$	R=0	R=1
S=0	0.01	0.8
S=1	0.8	0.95

$P(P = 1 \| G)$	G=0	G=1
	0.2	0.7

Blocking nodes

diverging edges	head-to-tail	converging edges
		
$a \not\perp\!\!\!\perp b$	$a \not\perp\!\!\!\perp b$	$\not\leftrightarrow$ $a \perp\!\!\!\perp b$
		
$\not\leftrightarrow$ $a \perp\!\!\!\perp b \mid c$	$\not\leftrightarrow$ $a \perp\!\!\!\perp b \mid c$	$a \not\perp\!\!\!\perp b \mid c$

The configuration with converging edges is called a v-structure

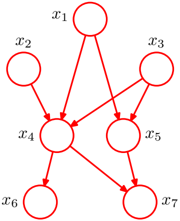
Factorization and Independence

- A factorization imposes independence statements

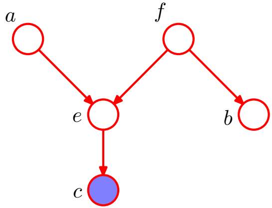
$$\forall x,\ p(x) = \prod_{j=1}^{p} p(x_j | x_{\Pi_j}) \quad \Leftrightarrow \quad \forall j,\ X_j \perp\!\!\!\perp X_{\{1,\dots,j-1\}\backslash\Pi_j} \mid X_{\Pi_j}$$

- Is it possible to read from the graph the (conditional) independence statements that hold given the factorization.

$$X_5 \overset{?}{\perp\!\!\!\perp} X_2 \mid X_4$$



d-separation



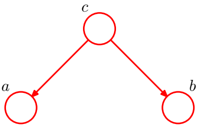
Theorem

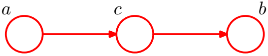
If A, B and C are three disjoint sets of node, the statement $X_A \perp\!\!\!\perp X_B | X_C$ holds if all paths joining A to B go through at least one *blocking node*. A node j is blocking a path

- if the edges of the paths are diverging/following and $j \in C$
- if the edges of the paths are converging (i.e. form a v-structure) and neither j nor any of its descendants is in C

Factorization et Independence II

- Several graphs can induce the same set of conditional independences .





$$p(c)p(a|c)p(b|c) = p(a)p(c|a)p(b|c)$$

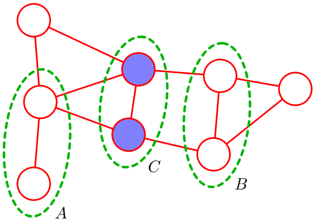
- Some combinations of conditional independences cannot be faithfully represented by a graphical model
  - Ex1: $X \sim \text{Ber}\frac{1}{2} \qquad Y \sim \text{Ber}\frac{1}{2} \qquad Z = X \oplus Y.$
  - Ex2: $X \perp\!\!\!\perp Y \mid Z = 1$ but $X \not\!\perp\!\!\!\perp Y \mid Z = 0$

Markov random field (MRF) or *Undirected graphical model*

Is it possible to associate to each graph a family of distribution so that conditional independence coincides exactly with the notion of separation in the graph?

Global Markov Property

$$X_A \perp\!\!\!\perp X_B \mid X_C \quad \Leftrightarrow C \text{ separates } A \text{ et } B$$



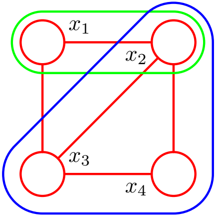
Gibbs distribution

Clique Set of nodes that are all connected to one another.

Potential function The potential $\psi_C(x_C) \geq 0$ is associated to clique C.

Gibbs distribution

$$p(x) = \frac{1}{Z} \prod_C \psi_C(x_C)$$



Partition function

$$Z = \sum_x \prod_C \psi_C(x_C)$$

Writing potential in exponential form $\psi_C(x_C) = \exp\{-E(x_C)\}$.
$E(x_C)$ is an *energy*.
This a *Boltzmann distribution*.

Example 1: Ising model

$X = (X_1, \ldots, X_d)$ is a collection of binary variables, whose joint probability distribution is

$$\begin{aligned} p(x_1, \ldots, x_d) &= \frac{1}{Z(\eta)} \exp\Big(\sum_{i \in V} \eta_i x_i + \sum_{\{i,j\} \in E} \eta_{ij} x_i x_j \Big) \\ &= \frac{1}{Z(\eta)} \prod_{i \in V} e^{\eta_i x_i} \prod_{\{i,j\} \in E} e^{\eta_{ij} x_i x_j} \\ &= \frac{1}{Z(\eta)} \prod_{i \in V} \psi_i(x_i) \prod_{\{i,j\} \in E} \psi_i(x_i, x_j) \end{aligned}$$

with $\psi_i(x_i) = e^{\eta_i x_i}$ and $\psi_{ij}(x_i, x_j) = e^{\eta_{ij} x_i x_j}$.

Example 2: Directed graphical model

Consider a distribution p that factorizes according to a directed graph $G = (V, E)$, then

$$\begin{aligned} p(x_1, \ldots, x_d) &= \prod_{i=1}^{d} p(x_i \mid x_{\pi_i}) \\ &= \prod_{i=1}^{d} \psi_{C_i}(x_{C_i}) \qquad \text{with} \quad C_i = \{i\} \cup \pi_i \end{aligned}$$

Consequence: A distribution that factorizes according to a directed model is a Gibbs distribution for the cliques $C_i = \{i\} \cup \pi_i$. As a consequence, it factorizes according to an undirected graph in which C_i are cliques.

Theorem of Hammersley and Clifford (1971)

A distribution p, which is such that $p(x) > 0$ for all x satisfies the *global Markov property* for graph G if and only if it is a Gibbs distribution associated with G.

- Gibbs distribution: $\mathcal{P}_G : \; p(x) = \frac{1}{Z} \prod_{C \in \mathcal{C}_G} \psi_C(x_C)$
- Global Markov property:

$$\mathcal{P}_M : \; X_A \perp\!\!\!\perp X_B \mid X_C \quad \text{si} \quad C \text{ separated } A \text{ and } B \text{ in } G$$

Theorem

We have $\quad \mathcal{P}_G \Rightarrow \mathcal{P}_M \quad$ and $\quad$ (HC): if $\forall x, \; p(x) > 0$, then $\mathcal{P}_M \Rightarrow \mathcal{P}_G$

Markov Blanket in an undirected graph

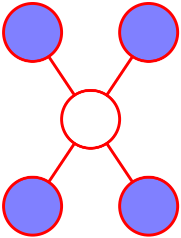
Definition

The Markov Blanket B of a node i is the smallest set of nodes B such that

$$X_i \perp\!\!\!\perp X_R \mid X_B, \qquad \text{with} \quad R = V \backslash (B \cup \{i\})$$

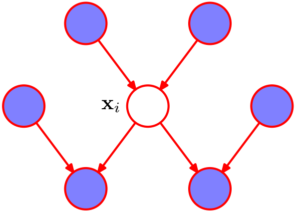
or equivalently such that

$$p(X_i \mid X_{-i}) = p(X_i \mid X_B).$$



Markov Blanket for a directed graph?

What is the Markov Blanket in a directed graph? By definition: the smallest set C of nodes such that conditionally on X_C, j is independent of all the other nodes in the graph?



Moralization

For a given oriented graphical model

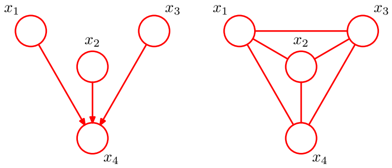
- is there an unoriented graphical model which is equivalent?
- is there a smallest unoriented graphical which contains the oriented graphical model?

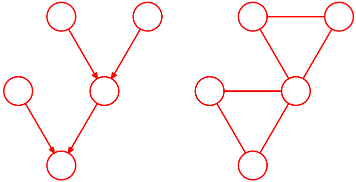
$$p(x) = \frac{1}{Z} \prod_C \psi_C(x_C) \quad \text{vs} \quad \prod_{j=1}^{M} p(x_j | x_{\Pi_j})$$

Moralization

Given a directed graph G, its moralized graph G_M is obtained by

1. For any node i, add undirected edges between all its parents
2. Remove the orientation of all the oriented edges





Proposition

If a probability distribution factorizes according to a directed graph G then it factorizes according to the undirected graph G_M.

Directed vs undirected trees

Definition: directed tree

A directed tree is a DAG such that each node has at most one parent

Remark: By definition a directed tree has no v-structure.

Moralizing trees

- What is the moralized graph for a directed tree?
- The corresponding undirected tree!

Proposition (Equivalence between directed and undirected tree)

A distribution factorizes according to a directed tree if and only if it factorizes according to its undirected version.

Corollary

All orientations of the edges of a tree that do not create v-structure are equivalent.

Operations on graphical models

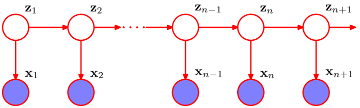
Probabilistic inference

Compute a marginal distribution $p(x_i)$ or a *conditional marginal* $p(x_i|x_1 = 3, x_7 = 0)$

Decoding (aka MAP Inference)

Finding what is the most probable configuration for the set of random variables?

$$\text{argmax}_z p(z|x)$$



Learning/ estimation in graphical models

Frequentist learning

The main *frequentist* learning principle for graphical model is the *maximum likelihood principle* of R. Fisher. Let $p(x;\boldsymbol{\theta}) = \frac{1}{Z(\boldsymbol{\theta})}\prod_C \psi(x_C,\theta_C)$, we would like to find

$$\text{argmax}_{\boldsymbol{\theta}} \prod_{i=1}^{n} p(x^{(i)};\boldsymbol{\theta}) \;=\; \text{argmax}_{\boldsymbol{\theta}} \frac{1}{Z(\boldsymbol{\theta})} \prod_{i=1}^{n}\prod_{C} \psi(x_C^{(i)},\theta_C)$$

Bayesian learning

Graphical models can also learn using *bayesian inference*.