Sequential Monte Carlo Methods For Probabilistic Graphical Models

Gautier Appert, Lionel Riou-Durand, Guillaume Salha

ENS Cachan

1 - Objectives

General context:

- Inference in statistical models involving a large number of latent variables is in general a difficult problem
- **Probabilistic Graphical Models** are a relevant and useful way to tackle this issue in many cases

In this project, we focus on **Sequential Monte Carlo methods** to solve **inference problems** in graphical models:

- We provide a review of the NIPS 2014 article of Naesseth, Lindsten and Schöm [1]
- We explain their new framework to make use of SMC methods in PGM
- We implement the method on Classical XY Model

2 - Graphical Models

We consider **undirected models** of the form:

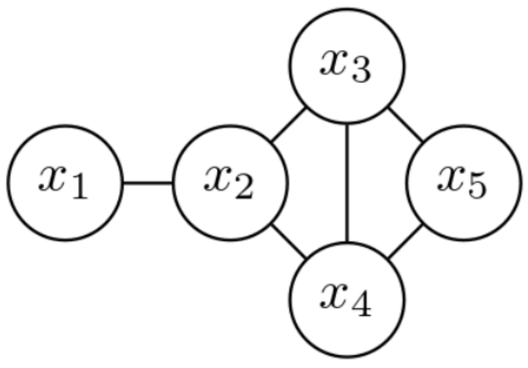
$$p(X_{\mathcal{V}}) = \frac{1}{Z} \prod_{c \in \mathcal{C}} \psi_c(X_c),$$

where the graph $\mathcal{G} = \{\mathcal{V}, \mathcal{E}\}$ has vertex set $\mathcal{V} = \{X_1, ..., X_{|\mathcal{V}|}\}$, edge set $\mathcal{E}$, cliques $\mathcal{C}$ and $Z = \sum_X \prod_{c \in \mathcal{C}} \psi_c(X_c)$ is the **partition function** (normalizing constant).

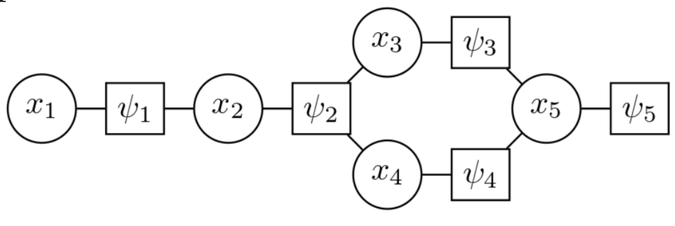
To make interactions between variables more explicit, [1] define a **factor graph** $\mathcal{F} = \{\mathcal{V}, \Psi, \mathcal{E}'\}$, associated to $\mathcal{G}$:

- Vertices: original r.v. + factors $\Psi = \{\psi_c : c \in \mathcal{C}\}$
- Edges: from variables to factors

For instance, the following graphical model:



could be associated to the following factor graph:



3 - Sequential Decomposition of Graphical Models

In order to use SMC methods for inference in graphical models, it is necessary to define a **sequence of target probability distributions** to approximate:

- We could intuitively want to use marginal distributions under $p(X_{\mathcal{V}})$. However, this is not an obligation ([1])
- The choice of target distribution is theoretically quite arbitrary as long as, at some final iteration, we recover $p(X_{\mathcal{V}})$
- In [1], they obtain such a sequence by doing a **sequential decomposition** of the graph

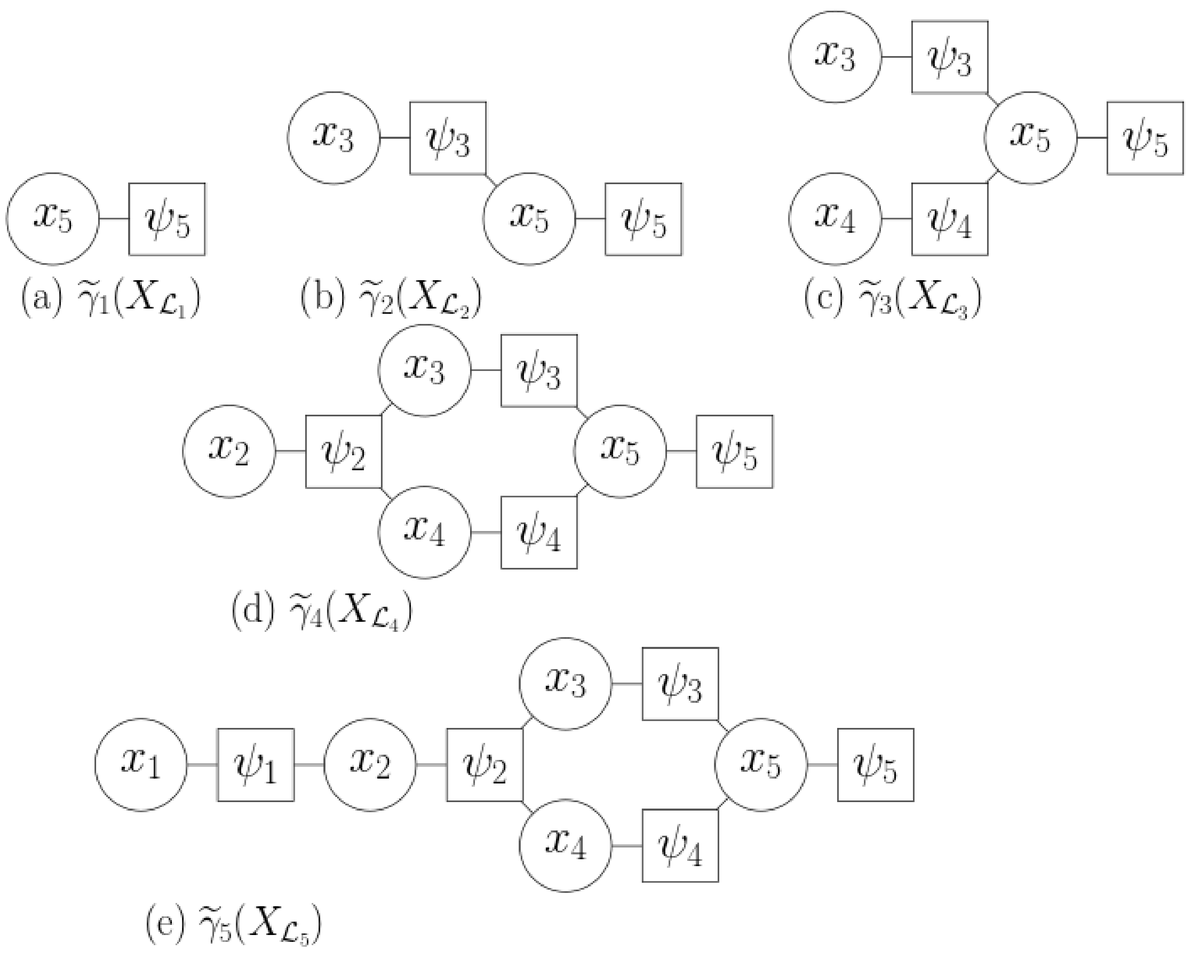
Sequential decomposition: Define a sequence of factors $\{\psi_k\}_{k=1}^K$ such that

$$\psi_k(X_{\mathcal{I}_k}) = \prod_{c \in \mathcal{C}_k} \psi_c(X_c),$$

where the different $\mathcal{C}_k \subset \mathcal{C}$ verify $\bigcup_{k=1}^K \mathcal{C}_k = \mathcal{C}$, and where $\mathcal{I}_k = \bigcup_{c \in \mathcal{C}_k} c \subseteq \{1, ..., |\mathcal{V}|\}$.
Therefore, the factorizarion of p can be rewritten as:

$$p(X_{\mathcal{V}}) = \frac{1}{Z} \prod_{k=1}^K \psi_k(X_{\mathcal{I}_k}).$$

We report a possible decomposition of the previous factor graph:



(a) $\tilde{\gamma}_1(X_{\mathcal{L}_1})$ (b) $\tilde{\gamma}_2(X_{\mathcal{L}_2})$ (c) $\tilde{\gamma}_3(X_{\mathcal{L}_3})$

(d) $\tilde{\gamma}_4(X_{\mathcal{L}_4})$

(e) $\tilde{\gamma}_5(X_{\mathcal{L}_5})$

[1] denote the auxiliary quantities by $\tilde{\gamma}_k(X_{\mathcal{L}_k}) := \prod_{l=1}^k \psi_l(X_{\mathcal{I}_l})$, with $\mathcal{L}_k := \cup_{l=1}^k \mathcal{I}_l$ and $\mathcal{L}_K = \mathcal{V}$. So:

$$p(X_{\mathcal{L}_K}) \propto \tilde{\gamma}_K(X_{\mathcal{L}_K}).$$

Therefore, we obtain the correct target distribution at time K.

The corresponding normalized probability distribution functions are given by $\bar{\gamma}_k(X_{\mathcal{L}_k}) = \tilde{\gamma}_k(X_{\mathcal{L}_k})/Z_k$ with $Z_k = \int \tilde{\gamma}_k(X_{\mathcal{L}_k}) dX_{\mathcal{L}_k}$.
An important requirement:

$$\int \tilde{\gamma}_k(X_{\mathcal{L}_k}) dX_{\mathcal{L}_k} < +\infty, \forall\, k \in \{1, ..., K\}$$

4 - Importance Sampling

Sequential Monte-Carlo methods relate to **Importance Sampling**, a well known Monte-Carlo method for:

- Approximating intractable integrals
- More generally, sample indirectly from a target distribution

Let π be an intractable density, $(X^{(1)}, \ldots, X^{(N)})$ be i.i.d. samples from a positive **proposal distribution** q from which we can sample. We define the importance weights as $w(x) := \frac{\gamma(x)}{q(x)}$.

The **normalized weights** are $W(X^{(i)}) := \frac{w(X^{(i)})}{\sum_{j=1}^N w(X^{(j)})}$. We derive using the LLN:

$$\hat{\mathcal{I}}_{\text{NIS}} := \sum_{i=1}^N h(X^{(i)}) W(X^{(i)}) \xrightarrow{\text{p.s}} \mathbb{E}_{X \sim \pi}[h(X)]$$

It also provides a **point-mass approximation** of the target distribution.
However, IS collapses when the dimension is too large as the variance of the weights grows exponentially. To tackle this issue, we describe the SMC sampler, in section 5.