Inference in probabilistic graphical models using sequential Monte Carlo

"Standard SMC methods using a non-standard construction of the intermediate target distributions"

Thomas Schön
Division of Systems and Control
Department of Information Technology
Uppsala University.

Email: thomas.schon@it.uu.se,
www: user.it.uu.se/~thosc112

Joint work with: Christian A. Naesseth (Linköping University) and Fredrik Lindsten (Uppsala University).
Performing inference in statistical models involving a large number of random variables and nonlinear interactions is a hard problem. Exploiting problem structure is vital. Probabilistic graphical models are a natural way to represent and make use of underlying structure.

Key idea: Introduce a sequential decomposition of the graphical model. The decomposition opens up for solving the inference problems using sequential Monte Carlo.

- Provides an approximation of the joint distribution
- and an unbiased estimate of the normalization constant.
A **probabilistic graphical model** (PGM) is a probabilistic model where a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ represents the conditional independency structure between random variables,

1. a set of **vertices** $\mathcal{V}$ (nodes) represents the random variables
2. a set of **edges** $\mathcal{E}$ containing elements $(i, j) \in \mathcal{E}$ connecting a pair of nodes $(i, j) \in \mathcal{V} \times \mathcal{V}$
Background – graphical models

The PGM factorizes according to its graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$

$$p(X) = \frac{1}{Z} \prod_{c \in \mathcal{C}} \psi_C(X_c),$$

where $X = \{x_1, \ldots, x_n\}$ and the normalization constant (partition function) is given by

$$Z = \int \prod_{c \in \mathcal{C}} \psi_C(X_c) dX.$$

Challenges:

1. Compute the joint distribution $p(X)$.
2. Compute the partition function. Relevant for:
   a) Likelihood-based learning of parameters in the PGM.
   b) Capacity calculations of a channel (information theory).
   c) Free energy of a system of objects (statistical mechanics).
Sequential Monte Carlo (SMC)

The distribution of interest $\pi(x)$ is called *target distribution*.

(Abstract) problem formulation: Sample from a sequence of probability distributions \( \{\pi_t(x_{1:t})\}_{t \geq 1} \) defined on a sequence of spaces of increasing dimension, where

\[
\pi_t(x_{1:t}) = \frac{\gamma_t(x_{1:t})}{Z_t},
\]

such that $\gamma_t(x_t): X^t \mapsto \mathbb{R}^+$ is known pointwise and $Z_t = \int \pi(x_{1:t})dx_{1:t}$ is often computationally challenging.

1. Approximate the normalizing constant $Z_t$.
2. Approximate $\pi_t(x_t)$ and compute integrals $\int \varphi(x_t)\pi_t(x_t)dx_t$.

Important question: How general is this formulation?
Sequential Monte Carlo (SMC)

The sequence of target distributions \( \{ \pi_t(x_{1:t}) \}_{t=1}^n \) can be constructed in many different ways.

The most basic construction arises from chain-structured graphs, such as the state space model (SSM).

\[
\pi_t(x_{1:t}) = \frac{\gamma_t(x_{1:t})}{p(y_{1:t})}
\]

\[
p(x_{1:t} \mid y_{1:t}) = \frac{p(x_{1:t}, y_{1:t})}{p(y_{1:t})}
\]

\[
Z_t = \int \pi_t(x_{1:t}) \, dx_{1:t} = p(y_{1:t})
\]
Sequential Monte Carlo (SMC)

The particle filter approximates $p(x_{1:t} \mid y_{1:t})$ for an SSM,

$$
x_{t+1} \mid x_t \sim f(x_{t+1} \mid x_t),
$$

$$
y_t \mid x_t \sim g(y_t \mid x_t),
$$

$$
x_1 \sim \mu(x_1),
$$

by maintaining an empirical distribution made up of $N$ samples (particles) $\{x_{1:t}^i\}_{i=1}^N$ and corresponding weights $\{w_{1:t}^i\}_{i=1}^N$

$$
\hat{p}(x_{1:t} \mid y_{1:t}) = \sum_{i=1}^N \frac{w_{1:t}^i}{\sum_{j=1}^N w_{1:t}^j} \delta_{x_{1:t}^i}(x_{1:t}).
$$
Example – indoor localization

Aim: Compute the position of a person moving around indoors using sensors (inertial, magnetometer and radio) located in an ID badge and a map, i.e. compute $p(x_t | y_{1:t})$. 
Example – indoor localization

Show movie

Sequential Monte Carlo – particle filter

SMC = sequential importance sampling + resampling

1. Propagation: \( x_t^i \sim f(x_t | x_{1:t-1}^{a_i}) \) and \( x_{1:t}^i = \{ x_{1:t-1}^{a_i}, x_t^i \} \).

2. Weighting: \( \bar{w}_t^i = W_t(x_t^i) = g(y_t | x_t^i) \).

3. Resampling: \( \mathbb{P}(a_t^i = j) = \bar{w}_t^j / \sum_l \bar{w}_{t-1}^l \).

The ancestor indices \( \{a_t^i\}_{i=1}^N \) are very useful auxiliary variables! They make the stochasticity of the resampling step explicit.
Key idea and our strategy

SMC methods are used to approximate a sequence of probability distributions on a sequence of spaces of increasing dimension.

**Key idea:**
1. Introduce a **sequential decomposition** of the PGM.
2. Each **subgraph** induces an intermediate target dist.
3. Apply SMC to the sequence of intermediate target dist.

Using an artificial sequence of intermediate target distributions for an SMC method is a powerful (quite possibly underutilized) idea.
Algorithm SMC for graphical models

1. **Initialize** ($k = 1$):
   (a) Draw $X^i_{\mathcal{L}_1} \sim r_1(\cdot)$.
   (b) Set $w^i_1 = W_1(X^i_{\mathcal{L}_1})$.

2. **For** $k = 2$ **to** $K$ **do**:
   (a) **Resampling**: Draw $a^i_k$, $\mathbb{P}(a^i_k = j) = \bar{w}^j_{k-1} / \sum_l \bar{w}_l^{k-1}$.
   (b) **Propagation**: Draw $\xi^i_k \sim r_k(\cdot | X^{a^i_k}_{\mathcal{L}_{k-1}})$, set $X^i_{\mathcal{L}_k} = X^{a^i_k}_{\mathcal{L}_{k-1}} \cup \xi^i_k$.
   (c) **Weighting**: Set $w^i_k = W_k(X^i_{\mathcal{L}_k})$.

$\mathcal{L}_k$ – index to the nodes in the $k^{th}$ intermediate target $\gamma_k(X_{\mathcal{L}_k})$.
$\xi^i_k$ – nodes added at step $k$.

Also provides an unbiased estimate of the **partition function**!


Thomas Schöen
Illustrating possible decompositions

Using a 2D lattice model from statistical physics. \( x \in (-\pi, \pi] \).

\[
p(X_V) \propto e^{-\beta H(X_V)}, \quad H(X_V) = - \sum_{(i,j) \in \mathcal{E}} J_{ij} \cos (x_i - x_j),
\]

The intermediate sequence of target distributions can be chosen

\[
\gamma_k(X_{\mathcal{L}_k}) \propto \gamma_{k-1}(X_{\mathcal{L}_{k-1}}) e^{\kappa(X_{\mathcal{L}_{k-1}})} \cos (x_k - \mu(X_{\mathcal{L}_{k-1}})).
\]

\( \mathcal{L}_k \) – index to the nodes in the \( k^{\text{th}} \) intermediate target \( \gamma_k(X_{\mathcal{L}_k}) \).
This example is borrowed from

Example 2 – 2D channel capacity

Problem: Compute the channel capacity for a 2D binary-input channel with the constraint that no two horizontally or vertically adjacent variables may be both be equal to 1.

\[
\begin{array}{ccc}
\ldots & \ldots & \ldots \\
\ldots & 0 & 1 & 0 & \ldots \\
\ldots & 0 & 0 & 1 & \ldots \\
\ldots & 0 & 1 & 0 & \ldots \\
\ldots & \ldots & \ldots & \ldots & \ldots \\
\end{array}
\]

Relevance: magnetic and optical storage solutions.

Model: Square lattice (undirected graphical model).

Noiseless capacity for an \( M \times M \) channel: \( C_M = \frac{1}{M^2} \log_2 Z \).
Example 2 – 2D channel capacity \((60 \times 60)\)

Our SMC method compared to the tree sampler by


implemented according to


Example 3 – spatio-temporal model

Detecting droughts in north America based on the measured precipitation.

Particle MCMC = SMC + MCMC

A systematic way of combining SMC and MCMC.
Builds on an extended target construction.

**Intuitively:** SMC is used as a high-dimensional proposal mechanism on the space of state trajectories $X^T$.

**A bit more precise:** Construct a Markov chain with $p(\theta, x_{1:T} \mid y_{1:T})$ (or one of its marginals) as its stationary distribution. Learning states and parameters.

Exact approximations

Pioneered by the work

Conclusions

The **sequential decomposition** opens up for SMC.

- Does **not** require a sequential structure in the model.
- Outputs an approximation of the full joint distribution.
- Outputs an unbiased estimate of the normalization constant.
- Can be used within MCMC in a plug-and-play manner.

Preliminary (for now) web site:

http://www.it.uu.se/conferences/smc2017
References to some of our work

SMC for graphical models


Fredrik Lindsten, Adam M. Johansen, Christian A. Naesseth, Bonnie Kirkpatrick, Thomas B. Schön, John Aston and Alexandre Bouchard-Côté. Divide-and-Conquer with Sequential Monte Carlo. Journal of Computational and Graphical Statistics (JCGS), 2016. (Accepted for publication)

Self-contained introduction to particle smoothing


Particle Gibbs with ancestor sampling construction


Financial support from the Swedish Research Council (VR) and the Swedish Foundation for Strategic research (SSF) is gratefully acknowledged.