Sequential Monte Carlo methods

Lecture 15 – General Sequential Monte Carlo

Thomas Schön, Uppsala University

2017-08-29

Outline – Lecture 15

Aim: Show how SMC can be used for a much wider class of problems than inference in state-space models.

Outline:

1. Summary of day 3
2. Examples of probabilistic models
3. General SMC formulation
4. Locally optimal proposals

Summary of day 3

Summary of day 3

Simulate a Markov chain which is designed in such a way that its stationary distribution coincides with the target distribution.

An MCMC sampler generates the Markov chain $\{x[m]\}_{m=1}^{M}$ by:

- **Initialize:** set $x[1]$ arbitrarily.
- **For** $m = 2$ **to** M**:** sample $x[m] \sim \kappa(x[m-1], x^\star)$.

$\kappa(x, x^\star)$ is a **Markov kernel** on $\mathcal{X}$, i.e. a conditional distribution for the next state $x^\star$ given the current state x.

Summary of day 3

Algorithm 1 Pseudo-marginal Metropolis Hastings

1. **Initialize ($m = 1$):** Set $\theta[1]$ and run a particle filter for $\hat{z}[1]$.
2. **For** $m = 2$ **to** M**, iterate:**
   a. Sample $\theta' \sim q(\theta \mid \theta[m-1])$.
   b. Sample $\hat{z}' \sim \psi(z \mid \theta', y_{1:T})$ (i.e. run a particle filter).
   c. With probability
   $$\alpha = \min\left(1, \frac{\hat{z}' p(\theta')}{\hat{z}[m-1]p(\theta[m-1])} \frac{q(\theta[m-1] \mid \theta')}{q(\theta' \mid \theta[m-1])}\right)$$
   set $\{\theta[m], \hat{z}[m]\} \leftarrow \{\theta', \hat{z}'\}$ (accept candidate sample) and with prob. $1 - \alpha$ set $\{\theta[m], \hat{z}[m]\} \leftarrow \{\theta[m-1], \hat{z}[m-1]\}$ (reject candidate sample).

Summary of day 3

Gibbs kernel: This procedure defines a Markov kernel $\kappa(x, x^\star)$ with stationary distribution $\pi(x)$.

Particle Gibbs kernel: A Markov kernel $\kappa_{N,\theta}(x_{0:T}, x^\star_{0:T})$ on $\mathcal{X}^{T+1}$.

Particle Gibbs: Run a particle filter, but at each time step

- sample only $N - 1$ particles in the standard way.
- set the Nth particle deterministically: $x_t^N = x_t$ and $a_t^N = N$.
- At final time $t = T$, output $x^\star_{0:T} = x^b_{0:T}$ with $b \sim \mathcal{C}(\{w_T^i\}_{i=1}^N)$

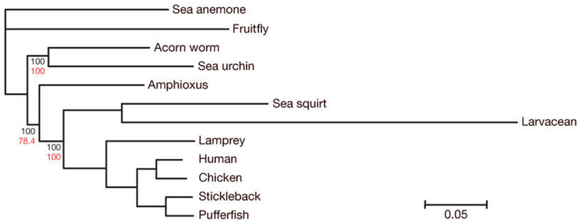
Summary of day 3

- The algorithm stochastically "maps" $x_{0:T}$ into $x^\star_{0:T}$.
- Implicitly defines a Markov kernel $\kappa_{N,\theta}(x_{0:T}, x^\star_{0:T})$ on $\mathcal{X}^{T+1}$ — **the particle Gibbs kernel**.

Examples of probabilistic models

Phylogenetic trees

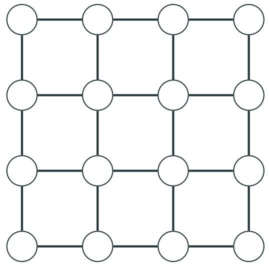
A phylogenetic (evolutionary) tree shows the inferred evolutionary relationships among various species based upon similarities and differences in their physical or genetic characteristics.

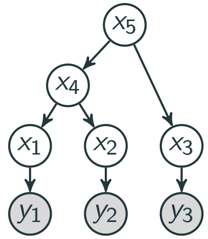


Probabilistic graphical models

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 $(i, j) \in \mathcal{V} \times \mathcal{V}$





Gaussian process state-space model

The Gaussian process (GP) is a **non-parametric** and **probabilistic** model for nonlinear functions.

Non-parametric means that it does not rely on any particular parametric functional form to be postulated.

$$
\begin{aligned}
X_t &= f(X_{t-1}) + V_t, \qquad && \text{s.t.} \quad f(X) \sim \mathcal{GP}(0, \kappa_{\eta,f}(x, x')), \\
Y_t &= g(X_t) + E_t, && \text{s.t.} \quad g(X) \sim \mathcal{GP}(0, \kappa_{\eta,g}(x, x')).
\end{aligned}
$$

The model functions f and g are assumed to be realizations from Gaussian process priors and $V_t \sim \mathcal{N}(0, Q)$, $E_t \sim \mathcal{N}(0, R)$.

Task: Compute the posterior $p(f, g, Q, R, \eta, x_{0:T} \mid y_{1:T})$.

General SMC formulation

Model specification

SMC can be used to approximate a **sequence** of probability distributions on a sequence of probability spaces of **increasing dimension**.

Let $\{\pi_k(x_{1:k})\}_{k\geq 1}$ be an arbitrary sequence of target distributions

$$\pi_k(x_{1:k}) = \frac{\widetilde{\pi}_k(x_{1:k})}{Z_k}$$

- The domain of x_k is $\mathcal{X}_k$, and $\mathcal{X}_{1:k} = \mathcal{X}_k \times \mathcal{X}_{1:k-1}$ for all k.
- $\widetilde{\pi}_k(x_{1:k})$ can be evaluated pointwise.
- The normalizing constant Z_k may be unknown.

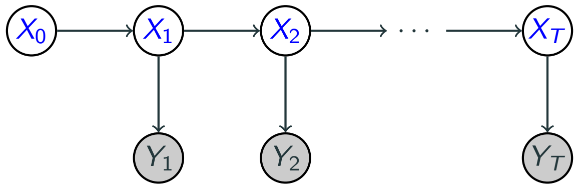
Common tasks:

1. Approximate the normalization constant Z_k.
2. Approximate $\pi_k(x_k)$ and compute $\int \phi(x_k)\pi_k(x_k)\mathrm{d}x_k$.

ex) State space model

The sequence of target distributions $\{\pi_k(x_{0:k})\}_{k=1}^n$ can be constructed in **many** different ways.

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



$$\overbrace{p(x_{0:t} \mid y_{1:t})}^{\pi_t(x_{0:t})} = \frac{\overbrace{p(x_{0:t}, y_{1:t})}^{\widetilde{\pi}_t(x_{0:t})}}{\underbrace{p(y_{1:t})}_{Z_t=\int \widetilde{\pi}(x_{0:t})\mathrm{d}x_{0:t}}}$$

General Sequential Monte Carlo

Sequential Monte Carlo approximates

$$\pi_k(x_{0:k}) \approx \sum_{i=1}^{N} w_k^i \delta_{x_{0:k}^i}(x_{0:k}).$$

The weighted particle populations $\{x_{0:k}^i, w_k^i\}_{i=1}^N$ are generated sequentially for $k = 1, 2, \ldots$

General Sequential Monte Carlo

Assume that we have obtained $\{x_{0:k-1}^i, w_{k-1}^i\}_{i=1}^N$

Resampling: Sample a_k^i with $\mathbb{P}(a_k^i = j) = \nu_{k-1}^j$, $j = 1, \ldots, N$.

Propagation: $x_k^i \sim q_k(x_k \mid x_{1:k-1}^{a_k^i})$ and $x_{0:k}^i = (x_{0:k-1}^{a_k^i}, x_k^i)$.

Weighting: $w_k^i \propto \dfrac{w_{k-1}^{a_k^i}}{\nu_{k-1}^{a_k^i}} \dfrac{\widetilde{\pi}_k(x_{0:k}^i)}{\widetilde{\pi}_{k-1}(x_{0:k-1}^{a_k^i}) q_k(x_k^i \mid x_{0:k-1}^{a_k^i})}$.

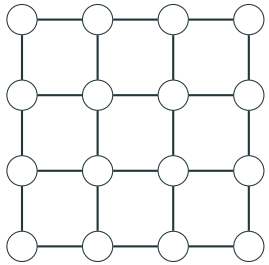
The result is a new weighted set of particles $\{x_{0:k}^i, w_k^i\}_{i=1}^N$.

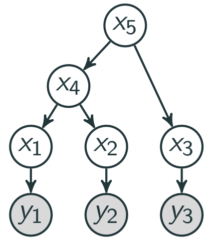
SMC for probabilistic graphical models

Recall – Probabilistic graphical models

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 $(i, j) \in \mathcal{V} \times \mathcal{V}$





Key idea

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

Key idea:

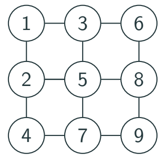
1. Introduce a **sequential decomposition** of the PGM.
2. Each **subgraph** induces an intermediate target dist.
3. Apply SMC to the sequence of intermediate target dist.

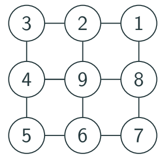
Using an artificial sequence of intermediate target distributions for an SMC method is a powerful (quite possibly underutilized) idea.

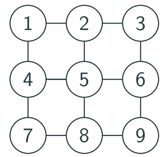
Key question: Exactly how do we define $\{\widetilde{\pi}_k(x_{1:k})\}_{k\geq 1}$?

ex) Illustrating possible graph decomposition

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







$$p(x_{\mathcal{V}}) \propto e^{-\beta H(x_{\mathcal{V}})}, \qquad H(x_{\mathcal{V}}) = -\sum_{(i,j)\in\mathcal{E}} J_{ij} \cos(x_i - x_j),$$

The **intermediate sequence** of target distributions can be chosen

$$\widetilde{\pi}_k(x_{\mathcal{L}_k}) \propto \widetilde{\pi}_{k-1}(x_{\mathcal{L}_{k-1}}) e^{\kappa(x_{\mathcal{L}_{k-1}})\cos(x_k - \mu(x_{\mathcal{L}_{k-1}}))}.$$

$\mathcal{L}_k$ – index to the nodes in the k^{th} intermediate target $\widetilde{\pi}_k(x_{\mathcal{L}_k})$.

SMC for graphical models – algorithm

Algorithm SMC for graphical models

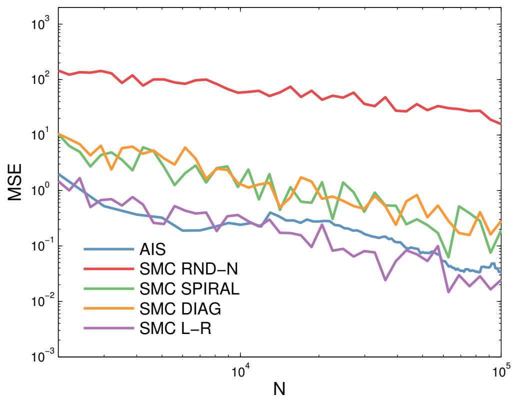
1. **Initialize ($k = 1$):**
   (a) Draw $x^i_{\mathcal{L}_1} \sim q_1(\cdot)$.
   (b) Set $w^i_1 = W_1(x^i_{\mathcal{L}_1})$.
2. **For** $k = 2$ **to** K **do:**
   (a) **Resampling:** Draw a^i_k, $\mathbb{P}(a^i_k = j) = \widetilde{w}^j_{k-1} / \sum_l \widetilde{w}^l_{k-1}$.
   (b) **Propagation:** Draw $\xi^i_k \sim q_k(\cdot | x^{a^i_k}_{\mathcal{L}_{k-1}})$, set $x^i_{\mathcal{L}_k} = x^{a^i_k}_{\mathcal{L}_{k-1}} \cup \xi^i_k$.
   (c) **Weighting:** Set $\widetilde{w}^i_k = W_k(x^i_{\mathcal{L}_k})$.
3. **End**

$\mathcal{L}_k$ – index to the nodes in the k^{th} intermediate target $\widetilde{\pi}_k(x_{\mathcal{L}_k})$.

ξ^i_k – nodes added at step k.

Also provides an unbiased estimate of the **normalizing constant**!

ex) Classical XY-model



This model is borrowed from

John M. Kosterlitz and David J. Thouless. **Ordering, metastability and phase transitions in two-dimensional systems.** *J. of Physics C: Solid State Physics*, 6(7):1181–1203, 1973.

What about stability?

Recall: for a state-space model we need exponential forgetting for the particle filter to be stable.

The same is true in the general case!
If there are **strong** and **long-ranging** dependencies among the variables $X_{1:k}$ under the distribution π_k, then the asymptotic variance of SMC may be exponential in k.

However,

- In many applications we *do* have fast enough forgetting (though, it can be difficult to verify theoretically)
- Even if this is not the case, SMC can give good results for moderate values of k

Further reading

SMC for phylogenetic trees:

Alexandre Bouchard-Côté. **Sequential Monte Carlo (SMC) for Bayesian phylogenetics.** *Bayesian phylogenetics: methods, algorithms, and applications*, 163-–186, 2014.

SMC for graphical models:

Christian A. Naesseth, Fredrik Lindsten and Thomas B. Schön. **Sequential Monte Carlo methods for graphical models.** *Advances in Neural Information Processing Systems (NIPS)*, Montreal, Canada, December, 2014.