Sequential Monte Carlo methods

Lecture 15 – General Sequential Monte Carlo

Fredrik Lindsten, Linköping University

2025-02-28

Outline – Lecture 15

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

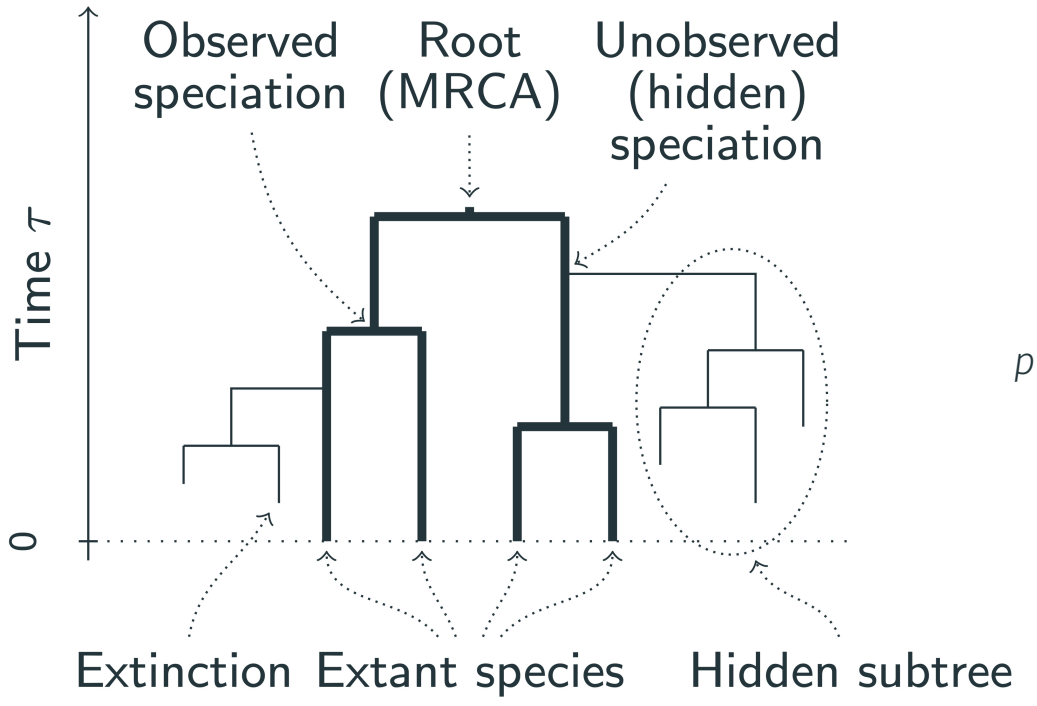
Outline:

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

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.

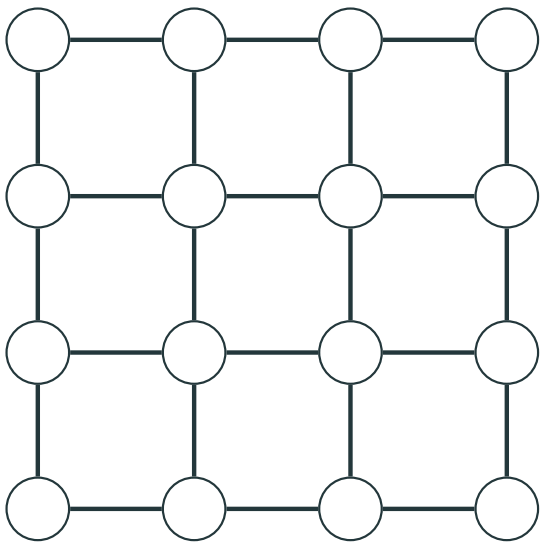


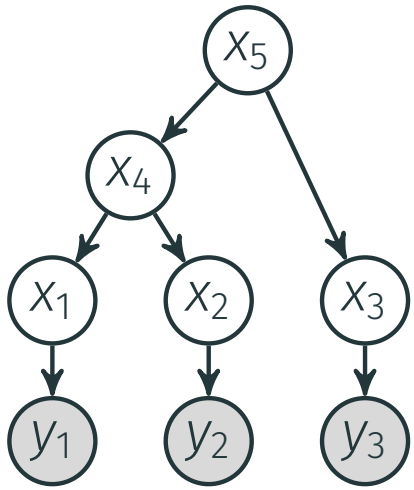
$$p(\mathcal{T}_r) = \begin{cases} 2\prod_{c \in \mathsf{Ch}(r)} p(\mathcal{T}_c) & \text{if } r \text{ is the root node,} \\ 2\lambda e^{-(\lambda+\mu)\Delta_r} \prod_{c \in \mathsf{Ch}(r)} p(\mathcal{T}_c) & \text{if } r \text{ is a speciation,} \\ \mu e^{-(\lambda+\mu)\Delta_r} & \text{if } r \text{ is an extinction,} \\ e^{-(\lambda+\mu)\Delta_r} & \text{if } r \text{ is an extant species,} \end{cases}$$

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 $V_t \sim \mathcal{N}(0, Q)$, $E_t \sim \mathcal{N}(0, R)$.

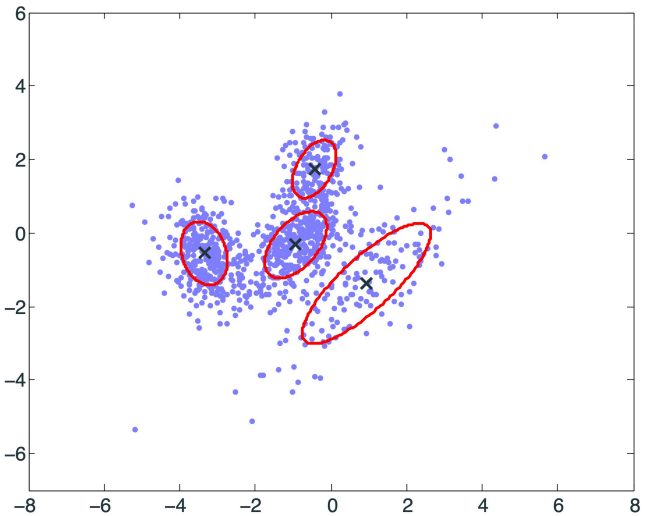
Task: Compute the posterior $p(f, g, Q, R, \eta, x_{0:T} \mid y_{1:T})$.

Dirichlet process mixtures models

The Dirichlet process is a Bayesian nonparametric model, which can be used to construct mixture models with an unknown and possibly infinite number of components.

Generative process:

$$
\begin{aligned}
p(x_{k+1} = j \mid x_{0:k}) &= \begin{cases} \frac{n_{k,j}}{k+\alpha} & \text{for } j = 1, \ldots, J_k, \\ \frac{\alpha}{k+\alpha} & \text{for } j = J_k + 1, \end{cases} \\
\theta_k &\sim F(\theta), \quad k = 0, 1, \ldots, \\
p(y_k \mid x_k, \{\theta_k\}_{k \geq 1}) &= G(y_k \mid \theta_{x_k}),
\end{aligned}
$$



Probabilistic models

Classes of probabilistic models to which SMC has been applied:

- State-space models
- Phylogenetics
- Generic PGMs
- Bayesian nonparametric models
- Probabilistic programming languages
- Diffusion generative models
- Autoregressive (language) models
- ...

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 $\{\pi_k(x_{0:k})\}_{k\geq 0}$ be an arbitrary sequence of target distributions

$$\pi_k(x_{0:k}) = \frac{\widetilde{\pi}_k(x_{0:k})}{Z_k}$$

Model specification

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

Let $\{\pi_k(x_{0:k})\}_{k\geq 0}$ be an arbitrary sequence of target distributions

$$\pi_k(x_{0:k}) = \frac{\widetilde{\pi}_k(x_{0:k})}{Z_k}$$

- The domain of x_k is $\mathcal{X}_k$, and $\mathcal{X}_{0:k} = \mathcal{X}_k \times \mathcal{X}_{0:k-1}$ for all k.

Model specification

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

Let $\{\pi_k(x_{0:k})\}_{k\geq 0}$ be an arbitrary sequence of target distributions

$$\pi_k(x_{0:k}) = \frac{\widetilde{\pi}_k(x_{0:k})}{Z_k}$$

- The domain of x_k is $\mathcal{X}_k$, and $\mathcal{X}_{0:k} = \mathcal{X}_k \times \mathcal{X}_{0:k-1}$ for all k.
- $\widetilde{\pi}_k(x_{0:k})$ can be evaluated pointwise.

Model specification

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

Let $\{\pi_k(x_{0:k})\}_{k\geq 0}$ be an arbitrary sequence of target distributions

$$\pi_k(x_{0:k}) = \frac{\widetilde{\pi}_k(x_{0:k})}{Z_k}$$

- The domain of x_k is $\mathcal{X}_k$, and $\mathcal{X}_{0:k} = \mathcal{X}_k \times \mathcal{X}_{0:k-1}$ for all k.
- $\widetilde{\pi}_k(x_{0:k})$ can be evaluated pointwise.
- The normalizing constant Z_k may be unknown.

Model specification

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

Let $\{\pi_k(x_{0:k})\}_{k\geq 0}$ be an arbitrary sequence of target distributions

$$\pi_k(x_{0:k}) = \frac{\widetilde{\pi}_k(x_{0:k})}{Z_k}$$

- The domain of x_k is $\mathcal{X}_k$, and $\mathcal{X}_{0:k} = \mathcal{X}_k \times \mathcal{X}_{0:k-1}$ for all k.
- $\widetilde{\pi}_k(x_{0:k})$ can be evaluated pointwise.
- The normalizing constant Z_k may be unknown.

Common tasks:

1. Approximate the normalization constant Z_k.
2. Approximate $\pi_k(x_{0:k})$ and compute $\int \phi(x_{0:k})\pi_k(x_{0:k})\mathrm{d}x_{0:k}$.

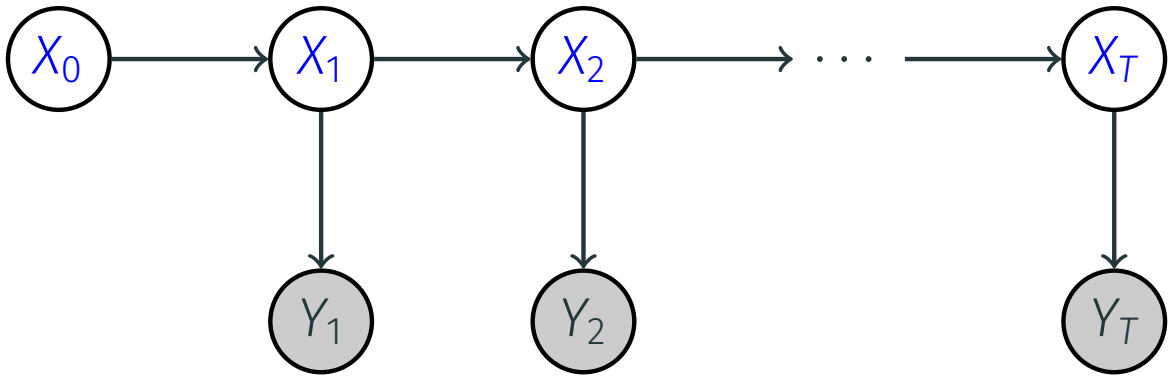
ex) State space model

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

ex) State space model

The sequence of target distributions $\{\pi_k(x_{0:k})\}_{k\geq 0}$ 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 $\{x^i_{0:k-1}, w^i_{k-1}\}_{i=1}^N$

Resampling: Sample a^i_k with $\mathbb{P}(a^i_k = j) = \nu^j_{k-1}$, $j = 1, \ldots, N$.

General Sequential Monte Carlo

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

Resampling: Sample a^i_k with $\mathbb{P}(a^i_k = j) = \nu^j_{k-1}$, $j = 1, \ldots, N$.

Propagation: $x^i_k \sim q_k(x_k \mid x^{a^i_k}_{0:k-1})$ and $x^i_{0:k} = (x^{a^i_k}_{0:k-1}, x^i_k)$.

General Sequential Monte Carlo

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

Resampling: Sample a^i_k with $\mathbb{P}(a^i_k = j) = \nu^j_{k-1}$, $j = 1, \ldots, N$.

Propagation: $x^i_k \sim q_k(x_k \mid x^{a^i_k}_{0:k-1})$ and $x^i_{0:k} = (x^{a^i_k}_{0:k-1}, x^i_k)$.

Weighting: $$w^i_k \propto \frac{w^{a^i_k}_{k-1}}{\nu^{a^i_k}_{k-1}} \frac{\widetilde{\pi}_k(x^i_{0:k})}{\widetilde{\pi}_{k-1}(x^{a^i_k}_{0:k-1}) q_k(x^i_k \mid x^{a^i_k}_{0:k-1})}.$$

General Sequential Monte Carlo

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

Resampling: Sample a^i_k with $\mathbb{P}(a^i_k = j) = \nu^j_{k-1}$, $j = 1, \ldots, N$.

Propagation: $x^i_k \sim q_k(x_k \mid x^{a^i_k}_{0:k-1})$ and $x^i_{0:k} = (x^{a^i_k}_{0:k-1}, x^i_k)$.

Weighting: $$w^i_k \propto \frac{w^{a^i_k}_{k-1}}{\nu^{a^i_k}_{k-1}} \frac{\widetilde{\pi}_k(x^i_{0:k})}{\widetilde{\pi}_{k-1}(x^{a^i_k}_{0:k-1}) q_k(x^i_k \mid x^{a^i_k}_{0:k-1})}.$$

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

Fully adapted general SMC

The analogue of the fully adapted particle filter in the general case is to use

Proposal: $$q_k(x_k \mid x^i_{0:k-1}) = \frac{\gamma_k((x^i_{0:k-1}, x_k))}{\nu_{k-1}(x^i_{0:k-1})\gamma_{k-1}(x^i_{0:k-1})}$$

Resampling weights: $$\nu^i_{k-1} \propto \nu_{k-1}(x^i_{0:k-1})w^i_{k-1}$$

where

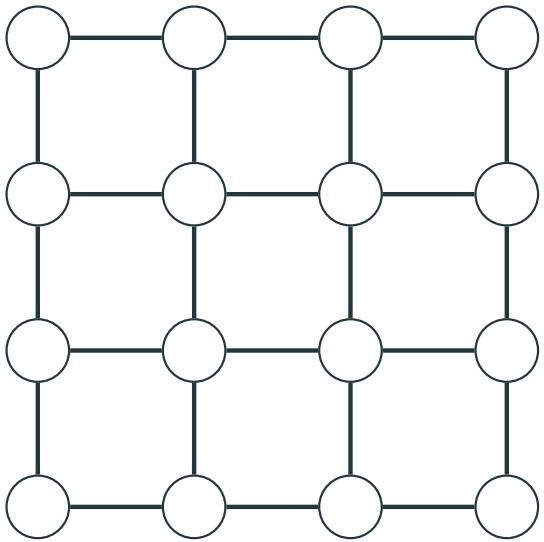
$$\nu_{k-1}(x_{0:k-1}) = \frac{\int \gamma_k(x_{0:k})dx_k}{\gamma_{k-1}(x_{0:k-1})}$$

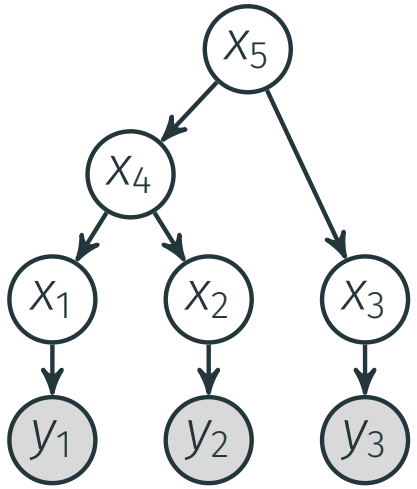
SMC for probabilistic graphical models

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

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.

Key idea

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

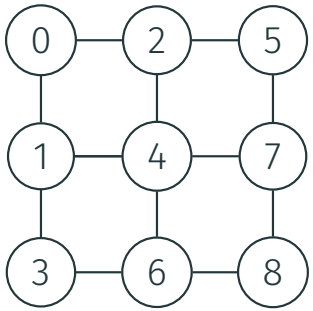
Key idea:

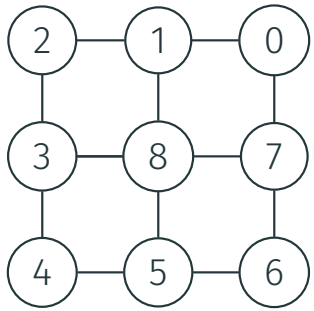
1. Introduce a sequential decomposition of the PGM.
2. Each subgraph induces an intermediate target dist.
3. Apply SMC to the sequence of intermediate target dist.

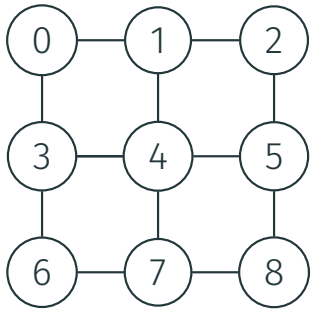
Using an artificial sequence of intermediate target distributions for an SMC method is a powerful (quite possibly underutilized) idea.

ex) Illustrating possible graph decomposition

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

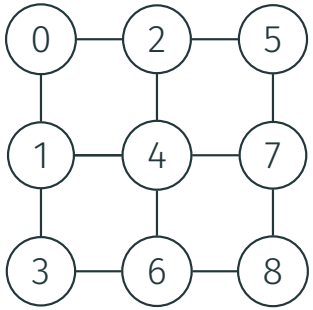


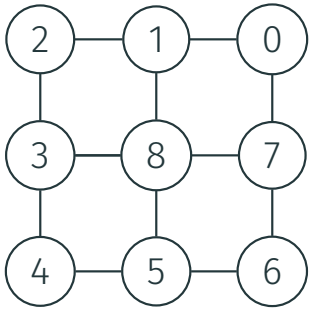


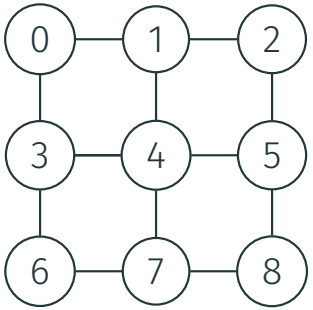


ex) Illustrating possible graph decomposition

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



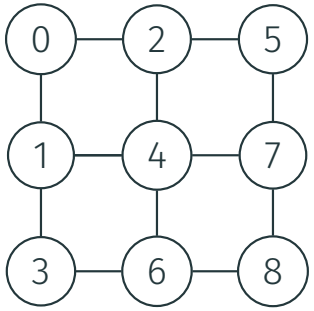


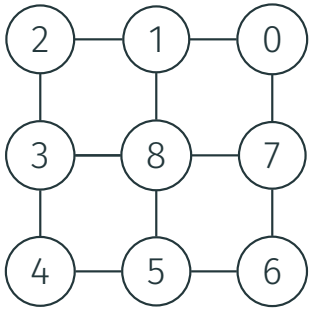


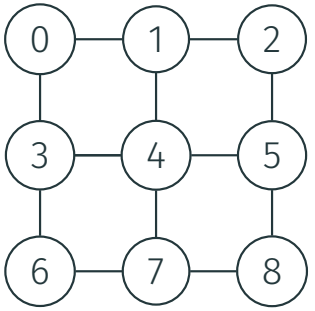
$$p(x_{0:K}) \propto e^{-\beta H(x_{0:K})}, \qquad H(x_{0:K}) = -\sum_{(i,j)\in\mathcal{E}} J_{ij} \cos(x_i - x_j),$$

ex) Illustrating possible graph decomposition

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





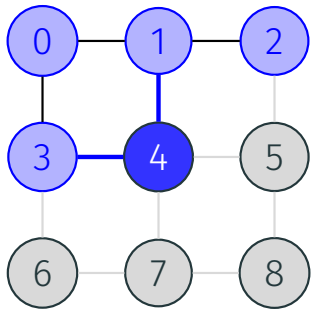


$$p(x_{0:K}) \propto e^{-\beta H(x_{0:K})}, \qquad H(x_{0:K}) = -\sum_{(i,j)\in\mathcal{E}} J_{ij} \cos(x_i - x_j),$$

The intermediate sequence of target distributions can be chosen

$$\widetilde{\pi}_k(x_{0:k}) \propto \exp\left(\beta \sum_{\substack{(i,j)\in\mathcal{E} \\ i,j\leq k}} J_{ij} \cos(x_i - x_j)\right)$$

ex) Illustrating possible graph decomposition



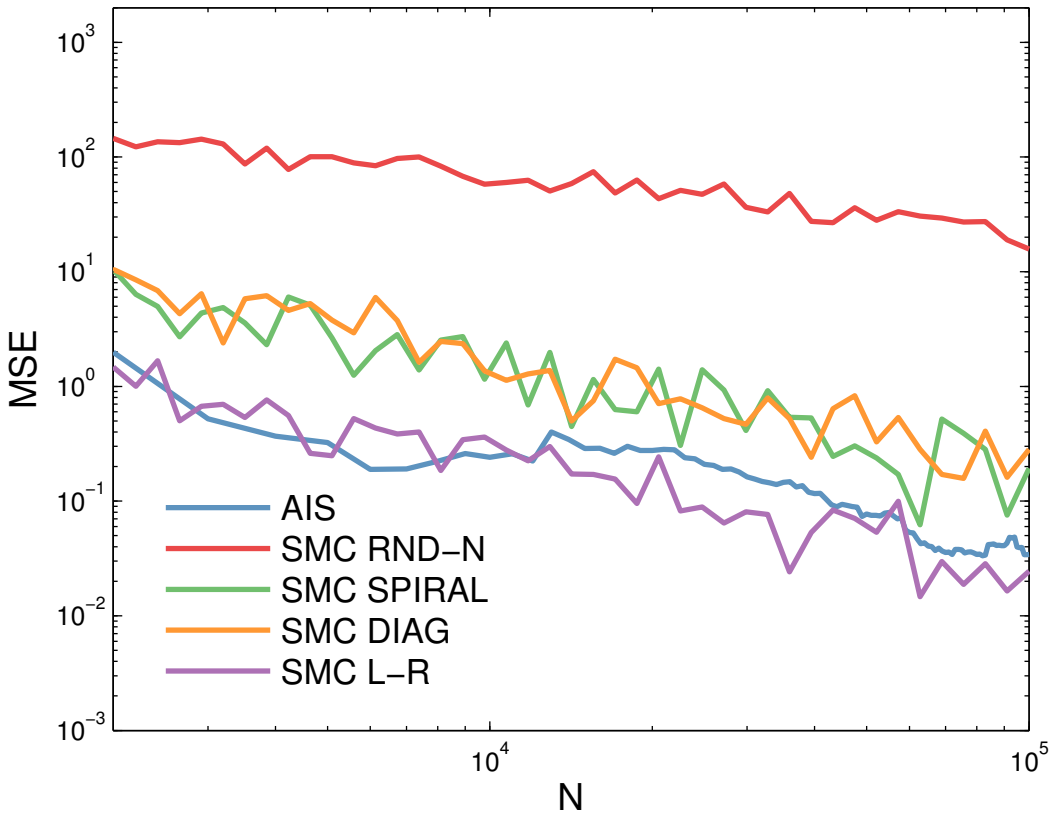
Ex. Sampling x_4 *conditionally on* $x_{0:3}$

Locally optimal proposal depends on

$$\frac{\widetilde{\pi}_k(x_{0:k})}{\widetilde{\pi}_{k-1}(x_{0:k-1})} \propto \exp\left(\beta \sum_{i \in \mathcal{N}_k} J_{ik} \cos\left(x_i - x_k\right)\right)$$

where $\mathcal{N}_k = \{i < k : (i,k) \in \mathcal{E}\}$.

ex) Classical XY-model



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

How to select the targets?

We want to sample from $p(x_{0:K}) \Rightarrow$ **Require:** $\pi_K(x_{0:K}) = p(x_{0:K})$, but...

...all intermediate targets $\{\pi_k(x_{0:k})\}_{k=0}^{K-1}$ are design choices!

How to select the targets?

We want to sample from $p(x_{0:K}) \Rightarrow$ **Require:** $\pi_K(x_{0:K}) = p(x_{0:K})$, but...

...all intermediate targets $\{\pi_k(x_{0:k})\}_{k=0}^{K-1}$ are design choices!

Idea:

$$\widetilde{\pi}_k^{\psi}(x_{0:k}) = \underbrace{\widetilde{\pi}_k(x_{0:k})}_{\text{From graph decomp.}} \overbrace{\psi_k(x_{0:k})}^{\text{Twisting funct.}}$$

- In theory, possible to select ψ_k to get exact samples from $p(x_{0:K})$ at iteration K.

How to select the targets?

We want to sample from $p(x_{0:K}) \Rightarrow$ **Require:** $\pi_K(x_{0:K}) = p(x_{0:K})$, but...

...all intermediate targets $\{\pi_k(x_{0:k})\}_{k=0}^{K-1}$ are design choices!

Idea:

$$\widetilde{\pi}_k^{\psi}(x_{0:k}) = \underbrace{\widetilde{\pi}_k(x_{0:k})}_{\text{From graph decomp.}} \overbrace{\psi_k(x_{0:k})}^{\text{Twisting funct.}}$$

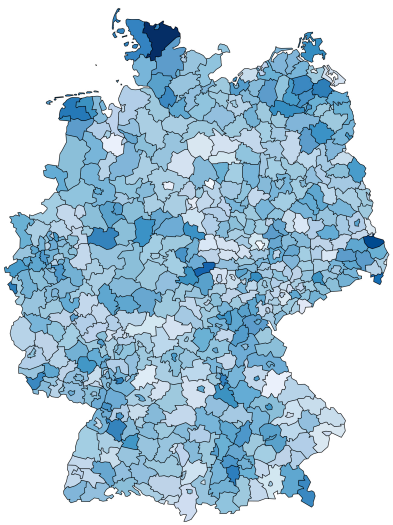
- In theory, possible to select ψ_k to get exact samples from $p(x_{0:K})$ at iteration K.
- Use deterministic inference method to approximate optimal ψ_k

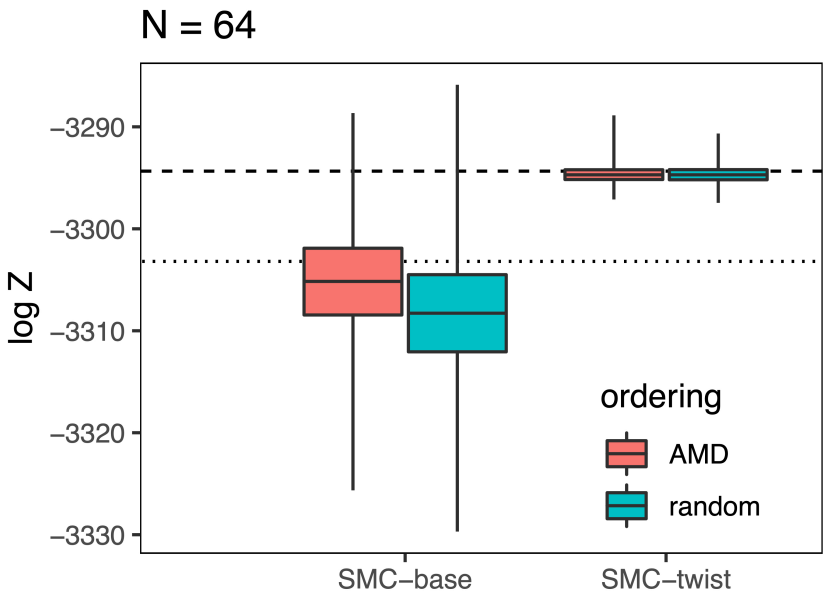
SMC works as a post-correction of the biases associated with the deterministic inference method of choice.

ex) CAR-Binomial

Model:

- Precision: $Q_{kk} = 0.1 \times \{|\mathrm{Ne}(t)| + 1\}$ and $Q_{kk'} = -0.1$ if $k \sim k'$
- $y_k \sim \mathrm{Binomial}(10, \mathrm{logit}^{-1}(x_k))$
- Spatial structure $\sim$ regions in Germany, $K = 544$





Fredrik Lindsten, Jouni Helske, and Matti Vihola. **Graphical model inference: Sequential Monte Carlo meets deterministic approximations.** *Advances in Neural Information Processing Systems (NeurIPS)*, Montreal, Canada, December, 2018.

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 $X_{0:k}$ 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

Summary

SMC can be used to approximate a sequence of probability distributions $\{\pi_k(x_{0:k})\}_{k\geq 0}$ (available up to an unknown normalization constant) on a sequence of spaces $\mathcal{X}_{0:k} = \mathcal{X}_k \times \mathcal{X}_{0:k-1}$ of increasing dimension.

Often we only care about (some marginal) of the final target $\pi_K(x_{0:K})$ in which case the intermediate targets $\{\pi_k(x_{0:k})\}_{k=0}^{K-1}$ are design choices.