Nested Sequential Monte Carlo Methods

Christian A. Naesseth, Fredrik Lindsten, Thomas B. Schön

Linköping University, Sweden
The University of Cambridge, UK
Uppsala University, Sweden

ICML 2015

Presented by: Qinliang Su

Aug. 17, 2016
Outline

1. Introduction
2. Review of Sequential Monte Carlo (SMC)
3. Nested SMC
4. Nesting of Nested SMC
5. Experiments
Introduction

Particle filters in high dimension
- Known to perform poorly in high (say, $d \gtrsim 10$) dimensions.
- Example: Spatio-temporal model:
  \[ g(y_t | x_t) = \prod_{k=1}^{d} g(y_{t,k} | x_{t,k}) \]
- Transition:
  \[ x_k | x_{k-1} \sim h(x_k | x_{k-1}) \]
- Measurement:
  \[ y_k | x_k \sim g(y_k | x_k) \]

Goal: at each time step $k$, use some samples to approximate the posterior

\[ p(x_{1:k} | y_{1:k}) \propto h(x_1)g(y_1 | x_1) \prod_{t=2}^{k} h(x_t | x_{t-1})g(y_t | x_t) \]

and then estimate the expectation $\mathbb{E}_p [f(x_{1:k})]$ as

\[ \mathbb{E}_{\hat{p}} [f(x_{1:k})] = \int f(x_{1:k})\hat{p}(x_{1:k})dx_{1:k} \]
This paper is interested in the settings:

1. $x_k$ is high-dimensional, i.e. $x_k \in \mathbb{R}^d$ with $d \gg 1$;
2. There are local dependency structure among $x_{1:k}$, both spatially and temporally

Two examples:
Outline

1. Introduction
2. Review of Sequential Monte Carlo (SMC)
3. Nested SMC
4. Nesting of Nested SMC
5. Experiments
High-level Descriptions of SMC

Procedures (at time step k):

i) Select one sequence from existing ones \( \{X_{1:k-1}^i\}_{i=1}^N \), denoted as \( X_{1:k-1}^i \)

ii) Draw a sample \( X_k^i \) from proposal distribution \( \bar{q}(x_k | X_{1:k-1}^i) \), and set \( X_{1:k}^i = (X_{1:k-1}^i, X_k^i) \) as the new sample

iii) Assign the new sample a weight \( W_k^i = \frac{p(x_{1:k}^i | y_{1:k})}{\bar{q}(X_{1:k}^i)} \), due to the mismatch between the proposal pdf and true pdf

\[
\{X_{1:k}^i, W_k^i\}_{i=1}^N \rightarrow \{\tilde{X}_{1:k}^i, 1/N\}_{i=1}^N.
\]

\[
\rightarrow \text{Propagation} \rightarrow \text{Weighting} \rightarrow \text{Resampling} \rightarrow \text{Weighting} \rightarrow \text{Resampling} \rightarrow \ldots
\]

With the pair \( \{X_{1:k}^i, W_k^i\}_{i=1}^N \), the posterior is approximated as

\[
p(x_{1:k}^i | y_{1:k}) \approx \sum_{i=1}^N \frac{W_k^i}{\sum_{i=1}^N W_k^i} \delta_{X_{1:k}^i} (x_{1:k}^i)
\]

The key is how to choose the proposal distribution
Bootstrap Particle Filter

Proposal pdf is chosen to be the transition pdf, i.e.,

$$\bar{q}(x_k|X_{1:k-1}^i) = h(x_k|X_{k-1}^i)$$

Under this proposal, the weight can be easily computed as

$$W_k^i = \frac{f(X_k^i|X_{k-1}^i)g(y_k|X_k^i)}{f(X_k^i|X_{k-1}^i)} = g(y_k|X_k^i)$$

Bootstrap PF performs poorly in high dimensions ($d > 10$)

- **Mismatch** between the proposal and target distributions
- Weight collapse, i.e. weights are dominated by only one weight

Despite of its simplicity, $h(x_t|x_{t-1})$ is a bad proposal distribution
The proposal pdf is chosen to adapt to the target distribution

Let $\pi_k(x_{1:k}) = \frac{1}{Z_{\pi_k}} \pi_k(x_{1:k})$ be the target pdf. The proposal pdf is designed as $\tilde{q}_k(x_k|x_{1:k-1}) = \frac{1}{Z_{q_k(x_{1:k-1})}} q_k(x_k|x_{1:k-1})$, where

$$q_k(x_k|x_{1:k-1}) = \frac{\pi_k(x_{1:k})}{\pi_{k-1}(x_{1:k-1})}, \quad [ = g(y_k|x_k) h(x_k|x_{k-1}) ]$$

Under this proposal pdf, the weight becomes

$$W_k^i = Z_{q_k(x_{1:k-1})}$$
Example: 2D MRF

Target pdf: \( \bar{\pi}(x_{1:k}) = \frac{1}{Z_{\pi_k}} \phi_1(x_1) \prod_{s=2}^{k} \phi_s(x_s) \psi_s(x_{s-1}, x_s) \)

Proposal pdf: \( q_k(x_k|x_{k-1}) = \phi_k(x_k) \psi_k(x_{k-1}, x_k) \)

Weight: \( Z_{q_k}(x_{k-1}) = \int \phi_k(x_k) \psi_k(x_{k-1}, x_k) dx_k \)
Fully Adapted SMC (3)

Algorithm 2:

- Select one sequence from \( \{X_{1:k-1}^j\}_{i=1}^N \) with probability proportional to \( \frac{Z_{q_k}(X_{1:k-1}^i)}{\sum_{i=1}^N Z_{q_k}(X_{1:k-1}^i)} \), denoted as \( X_{1:k-1}^j \);

- Draw \( X_k^j \) from \( \bar{q}_k(\cdot|X_{1:k-1}^j) \) and let \( X_{1:k}^j = (X_{1:k-1}^j, X_k^j) \)

Repeat above algorithm \( N \) times, we obtain samples \( \{X_{1:k}^j\}_{i=1}^N \), and obtain

\[
\bar{\pi}_k(x_{1:k}) \approx \frac{1}{N} \sum_{i=1}^N \delta_{X_{1:k}^i}(x_{1:k})
\]

However, exact computation of \( Z_{q_k} \) and sampling from \( \bar{q}_k(\cdot|X_{1:k-1}^j) \) are often impossible in practice.
Relaxing the exact computation and sampling requirements in fully adapted SMC......

**Definition 1 (Properly weighted sample)**

Let $\bar{q}(x) = \frac{1}{Z_q} q(x)$. A (random) pair $(X, W) \in \mathbb{R} \times \mathbb{R}_+$ is properly weighted w.r.t. $q(\cdot)$ if $\mathbb{E}_{(X,W)}[f(X)W] = Z_q \mathbb{E}_q[f(x)]$ for all measurable functions $f(x)$.

The exact pair $(X, W)$ with $X \sim \bar{q}(x)$ and $W = Z_q$ is a special case of properly weighted samples.
(A1) Let $Q$ be a class, and let $q = Q(q, M)$. Assume that:

i) The construction of $q$ returns a member variable $\hat{Z}_q = q.\text{GetZ}();$

ii) $Q$ has a member function $\text{Simulate}(\cdot)$ which returns a (possibly random) variable $X = q.\text{Simulate}()$

iii) $(X, \hat{Z}_q)$ is properly weighted w.r.t. $q()$
Nested SMC (3)

Replace the exact $Z_q$ and $X$ in fully adapted SMC with $q$.GetZ() and $q$.Simulate()

**Algorithm 3:**

- Initialize $q^i = Q(q_k(\cdot|X^i_{1:k-1}), M)$ for $i = 1, 2, \cdots, N$
- Set $\hat{Z}_{q_k}^i = q^i$.GetZ() for $i = 1, 2, \cdots, N$
- Repeat $N$ times
  - Select one element from $\{1, 2, \cdots s, \cdots, N\}$ with probabilities $\frac{\hat{Z}_s^{q_k}}{\sum_{s=1}^{N} \hat{Z}_s^{q_k}}$; denote the selected index as $j$
  - Draw $X_k^i = q^j$.Simulate() let $X_{1:k}^i = (X_{1:k-1}^i, X_k^i)$
Nested SMC (4)

**Theorem 1**

Assume $Q$ satisfies condition (A1). Then, the generated samples from nested SMC satisfies

$$N^{1/2} \left( \frac{1}{N} \sum_{i=1}^{N} f(X_{1:k}^i) - \bar{\pi}_k(f) \right) \xrightarrow{D} \mathcal{N}(0, \Sigma_k^M(f)),$$

where $\xrightarrow{D}$ means converges in distribution.

As long as $(q.\text{GetZ}, q.\text{Simulate}())$ is properly weighted, the expectation estimated from nested SMC converges to the exact expectation $\bar{\pi}_k(f)$ as $N$ increases.
Outline

1. Introduction
2. Review of Sequential Monte Carlo (SMC)
3. Nested SMC
4. Nesting of Nested SMC
5. Experiments
Nesting of Nested SMC (1)

\[ q_t(x_t | x_{t-1}) = \phi_t(x_t) \psi_t(x_{t-1}, x_t) \]

\[ Z_{\pi_k} \text{ can be estimated as: } \hat{Z}_{\pi_k} = \hat{Z}_{\pi_{k-1}} \times \left\{ \frac{1}{N} \sum_{i=1}^{N} \hat{Z}_{q_k}^i \right\}, \text{ where } \hat{Z}_{q_k}^i = q^i \cdot \text{GetZ}(). \]

**Theorem 2**

The pair \((X^i_{1:k}, \hat{Z}_{\pi_k}^i)\) is properly weighted w.r.t. \(\pi_k(\cdot)\), in which \(X^i_{1:k}\) is drawn with Algorithm 3.

**Implication:** using nested SMC, properly weighted samples w.r.t. 2D MRF \(\pi_k(\cdot)\) can be obtained from the properly weighted samples w.r.t. 1D MRF \(q_k(\cdot)\).
Nesting of Nested SMC (2)

(q^i. Simulate, q^i. GetZ) is properly weighted w.r.t. 1D MRF \( q(\cdot) \)

\[ \Downarrow \]

(\( X^i_{1:k}, \hat{Z}^i_{\pi_k} \)) is properly weighted w.r.t. 2D MRF \( \pi(\cdot) \)

(\( X^i_{1:k}, \hat{Z}^i_{\pi_k} \)) is properly weighted w.r.t. 2D MRF \( \pi(\cdot) \)

\[ \Downarrow \]

Draw samples from 3D MRF

- **Conclusion**: One nested SMC sampler can be used as the proposal distribution for another nested SMC targeting at higher dimensional distributions
Outline

1. Introduction
2. Review of Sequential Monte Carlo (SMC)
3. Nested SMC
4. Nesting of Nested SMC
5. Experiments
1) Gaussian State Space Model

**Figure:** Gaussian state space model in form of 2D MRF of size $d \times t$

The transition and measurement pdfs are all Gaussian

Two-level Nested SMC
Experiments (2)

Figure: Median effective sample size (ESS) and 15% ~ 85% percentiles. \( N = 500 \) and \( M = 2d \) with 100 independent runs.

\[
\text{ESS}(x_{k,\ell}) \triangleq \left( \mathbb{E} \left[ \frac{(\hat{x}_{k,\ell} - \mu_{k,\ell})^2}{\sigma_{k,\ell}^2} \right] \right)^{-1}
\]
2) Non-Gaussian State Space Model

- The transition pdf $p(x_k|x_{k-1})$ is Gaussian mixture
- The measurement pdf $p(y_k|x_k)$ is t-distribution

**Figure**: Median ESS and 15% ~ 85% percentiles.
3) Spatio-Temporal Model-Drought Detection

- Hidden states: 0 (normal) or 1 (drought) at different locations and years
- Measurements: precipitation
Experiments (5)

# of drought locations of North America in 1939

Estimate of $p(x_{k,i} = 1)$ for locations of North America in 1939