An Introduction to Particle Methods (a.k.a. Sequential Monte Carlo) for Filtering and Smoothing

Olivier Cappé

LTCI, Telecom ParisTech & CNRS
http://perso.telecom-paristech.fr/~cappe/

SIMINOLE Project, Oct. 2010

Hidden Markov Model (HMM)

The Hidden State Process $\{X_k\}_{k\geq 0}$ is a Markov chain with initial probability density function (pdf) $t_0(x)$ and transition density function $t(x, x')$ such that*

$$p(x_{0:k}) = t_0(x_0) \prod_{l=0}^{k-1} t(x_l, x_{l+1}) \ .$$

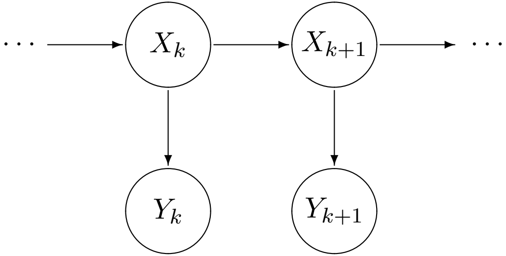
The Observed Process $\{Y_k\}_{k\geq 0}$ is such that the conditional joint density of $y_{0:k}$ given $x_{0:k}$ has the conditional independence (product) form

$$p(y_{0:k}|x_{0:k}) = \prod_{l=0}^{k} \ell(x_l, y_l) \ .$$

*$x_{0:k}$ denotes the collection $x_0, \ldots, x_k$.

Graphical Representation of the Dependence Structure

The HMM can be represented pictorially by a Bayesian network which depicts the conditional independence relations:



State-Space Form

Here the model is described in a functional form:

$$X_{k+1} = a(X_k, U_k) \; ,$$
$$Y_k = b(X_k, V_k) \; ,$$

where $\{U_k\}_{k\geq 0}$ and $\{V_k\}_{k\geq 0}$ are mutually independent i.i.d. sequences of random variables (also independent of X_0).

State-Space Form

Here the model is described in a functional form:

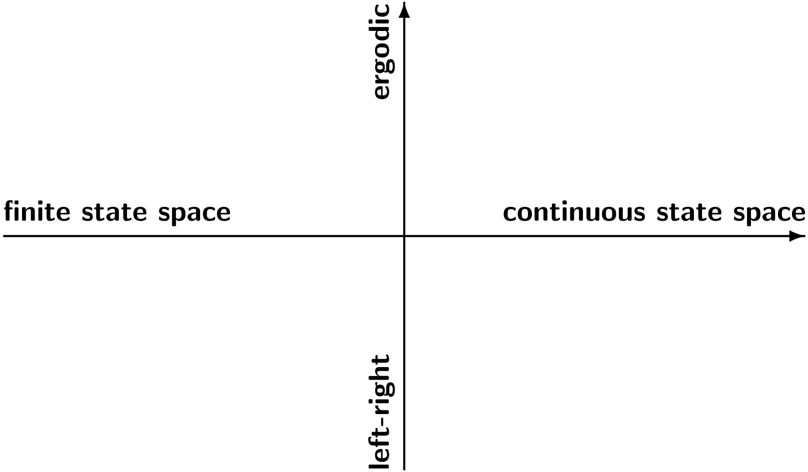
$$X_{k+1} = a(X_k, U_k) \ ,$$
$$Y_k = b(X_k, V_k) \ ,$$

where $\{U_k\}_{k \geq 0}$ and $\{V_k\}_{k \geq 0}$ are mutually independent i.i.d. sequences of random variables (also independent of X_0).

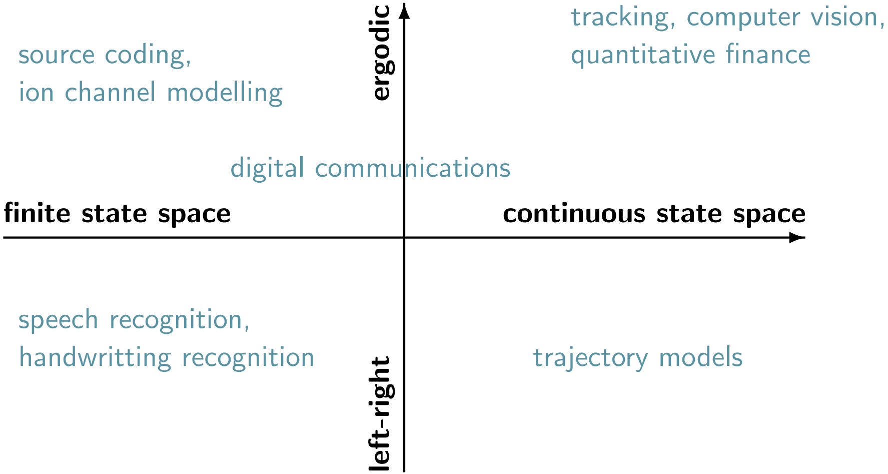
Remark

The term *state-space model* often refers to the case where a and b are linear functions of their arguments (and $\{U_k\}$, $\{V_k\}$, X_0 are jointly Gaussian).
Likewise, the term *HMM* is sometimes used (not in this talk!) more restrictively for the case where X is a finite set.

HMM Examples



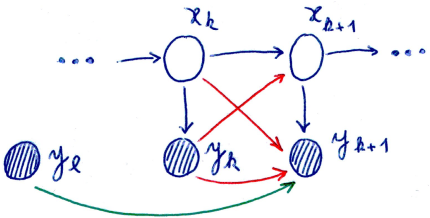
HMM Examples



Beyond HMMs

For sequential Monte Carlo methods, the key point is the structure of the joint conditional $p(x_{0:k}|y_{0:k})$. The methods described in this talk directly apply in cases where the joint conditional may be factored as

$$p(x_{0:k}|y_{0:k}) = p(x_0|y_0) \prod_{l=0}^{k-1} p(x_{l+1}|x_l, y_{0:l+1})$$



Tasks of interest for HMMs

State Inference: How to make probabilistic statements on the state sequence given the model and the observations?

- Filtering: $\pi_{k|k}(x_k) = p(x_k|Y_{0:k})$
- Prediction: $\pi_{k+1|k}(x_{k+1}) = p(x_{k+1}|Y_{0:k})$
- Smoothing: $\pi_{0:k|k}(x_{0:k}) = p(x_{0:k}|Y_{0:k})$
  (*fixed-interval*: $\pi_{l|k}$ for $l = 0, \ldots, k$; *fixed-lag*: $\pi_{k|k+\Delta}$ for $k = 0, \ldots$)

Parameter Inference: How to tune the model parameters based on the observations?

Recursive Structure of the Joint Smoothing Density

By Bayes' rule

$$
\begin{aligned}
&\pi_{0:k+1|k+1}(x_{0:k+1}) \\
&\quad = (\mathrm{L}_{k+1}(Y_{0:k+1}))^{-1} \, t_0(x_0) \prod_{l=0}^{k} t(x_l, x_{l+1}) \prod_{l=0}^{k+1} \ell(x_l, Y_l) \\
&\quad = \left(\frac{\mathrm{L}_{k+1}(Y_{0:k+1})}{\mathrm{L}_k(Y_{0:k})} \right)^{-1} \pi_{0:k|k}(x_{0:k}) \, t(x_k, x_{k+1}) \ell(x_{k+1}, Y_{k+1}) \; ,
\end{aligned}
$$

where the normalization constants L_k, i.e., the likelihood of the observations, is usually not computable.

The Joint Smoothing Recursion

$$\pi_{0:k+1|k+1}(x_{0:k+1}) = \left(\frac{\mathrm{L}_{k+1}}{\mathrm{L}_k}\right)^{-1} \pi_{0:k|k}(x_{0:k})\, t(x_k, x_{k+1})\ell(x_{k+1}, Y_{k+1})$$

The marginal recursion may be decomposed in two steps:

Prediction

$$\pi_{k+1|k}(x_{k+1}) = \int \pi_{k|k}(x_k) t(x_k, x_{k+1}) dx_k$$

Filtering

$$\pi_{k+1|k+1}(x_{k+1}) = \left(\frac{\mathrm{L}_{k+1}}{\mathrm{L}_k}\right)^{-1} \pi_{k+1|k}(x_{k+1})\ell(x_{k+1}, Y_{k+1})$$

Exact Implementation of the Filtering and Smoothing Recursions

When X is finite (Baum *et al.*, 1970) The computational cost of filtering is $|\mathsf{X}|^2$ per time index.

In linear Gaussian state-space models (Kalman & Bucy, 1961) The filtering and prediction recursion is implemented by the *Kalman filter* ($\mathrm{L}_{k+1}/\mathrm{L}_k$ is interpreted as the likelihood of the $(k+1)$-th innovation).

Such *finite dimensional filters* exist only in very specific models (see, e.g., Runggaldier & Spizzichino, 2001).

The Finite Case

Forward (Filtering) – Backward (Smoothing)

Forward For $k = 0$ up to $n-1$,

$$\pi_{k+1|k+1}(x_{k+1}) = \frac{\ell(x_{k+1}, Y_{k+1}) \sum_{x_k} \pi_{k|k}(x_k) t(x_k, x_{k+1})}{\sum_{x'} \ell(x', Y_{k+1}) \sum_{x} \pi_{k|k}(x) t(x, x')}$$

Backward For $k = n-1$ down to 0,

$$\pi_{k|n}(x_k) = \sum_{x_{k+1}} b_k(x_k|x_{k+1}) \pi_{k+1|n}(x_{k+1})$$

where

$$b_k(x_k|x_{k+1}) = \frac{\pi_{k|k}(x_k) t(x_k, x_{k+1})}{\sum_x \pi_{k|k}(x) t(x, x_{k+1})}$$
$$= \mathrm{P}(X_k = x_k | X_{k+1} = x_{k+1}, Y_{0:k})$$

Approximate Implementations of the Filtering and Smoothing Recursions

- EKF (Extended Kalman Filter) Linearization-based approach (for non-linear Gaussian state space models);
- UKF (Unscented Kalman Filter, Julier & Uhlmann, 1997) Point-based approach;
- and more Gaussian or Assumed Density Filters (ADF) (Wu, Hu, Xu & Hu, 2006).
- Variational Methods (e.g., Valpola & Karhunen, 2002) Based on parametric density approximation arguments.
- Exact Suboptimal Filters In particular, Kalman filter viewed as minimum mean square error linear filtering.

Sequential Monte Carlo (SMC)

- Sequential Monte Carlo (sometimes called *particle filtering*) is a method which uses pseudo-random simulations to produce successive populations of "particles" $X_k^{1:n}$ and associated weights $W_k^{1:n}$ such that

$$\sum_{i=1}^{n} W_k^i f(X_k^i) \approx \int f(x)\pi_{k|k}(x)dx \ ,$$

  for all functions f of interest.
- The SMC process is sequential in the sense that given $X_k^{1:n}$, $W_k^{1:n}$ and the observations $Y_{0:k+1}$, $X_{k+1}^{1:n}$ and $W_{k+1}^{1:n}$ are conditionally independent of previous populations of particles.
- SMC is based on importance sampling and resampling.

Self-Normalized Importance Sampling, or IS (Hammersley & Handscomb, 1964)

IS is a weighted form of Monte Carlo approximation, in which expectations under the target pdf π

$$\pi(f) = \mathrm{E}_\pi[f(X)]$$

are estimated as

$$\hat{\pi}_q^n(f) = \sum_{i=1}^{n} \underbrace{\frac{\omega^i}{\sum_{j=1}^n \omega^j}}_{W^i} f(X^i) = \frac{\frac{1}{n}\sum_{i=1}^n \omega^i f(X^i)}{\frac{1}{n}\sum_{j=1}^n \omega^j} ,$$

where

- $X^i \sim$ iid q, where q is an instrumental pdf
- $\omega^i = \frac{\pi}{q}(X^i)$.

This form of IS (sometimes also called Bayesian IS) does not necessitate that π be properly normalized.

Performance of IS

Assuming that $\mathrm{E}_\pi[\frac{\pi}{q}(X)(1 + f^2(X))] < \infty$, $\hat{\pi}_q^n(f)$ is consistent and asymptotically normal, with asymptotic variance given by

$$v_q(f) = \mathrm{E}_\pi\left[\frac{\pi}{q}(X)\left(f(X) - \pi(f)\right)^2\right] .$$

The asymptotic variance can be estimated from the IS sample by

$$\hat{v}_q^n(f) = n\sum_{i=1}^{n}(W^i)^2\{f(X^i) - \hat{\pi}_q^n(f)\}^2 ,$$

where $W^i = \omega^i / \sum_{j=1}^n \omega^j$ are the normalized weights.

Back to the Filtering and Smoothing Problem

How to estimate expectations under the posterior $\pi_{0:k|k}(x_{0:k}) = p(x_{0:k}|Y_{0:k})$ in the model

$$p(x_{0:k}) = t_0(x_0) \prod_{l=0}^{k-1} t(x_l, x_{l+1}) \ ,$$

$$p(y_{0:k}|x_{0:k}) = \prod_{l=0}^{k} \ell(x_l, y_l) \ ,$$

using a sequential algorithm ?

Sequential Smoothing through IS, or SIS (Handschin & Mayne, 1969-1970)

1. Propose n independent particle trajectories $\{X^i_{0:k+1}\}^{1\leq i\leq n}$ under a Markovian scheme such that

$$p(x_{0:k+1}) = \rho_{0:k+1}(x_{0:k+1}) = q_0(x_0)\prod_{l=1}^{k} q_l(x_l, x_{l+1}) \ .$$

Sequential Smoothing through IS, or SIS (Handschin & Mayne, 1969-1970)

1. Propose n independent particle trajectories $\{X^i_{0:k+1}\}^{1 \leq i \leq n}$ under a Markovian scheme such that

$$p(x_{0:k+1}) = \rho_{0:k+1}(x_{0:k+1}) = q_0(x_0) \prod_{l=1}^{k} q_l(x_l, x_{l+1}) \; .$$

2. Compute importance weights sequentially:

$$\omega^i_{k+1} = \frac{\pi_{0:k+1|k+1}(X^i_{0:k+1})}{\rho_{0:k+1}(X^i_{0:k+1})} = \omega^i_k \times \frac{t(X^i_k, X^i_{k+1})\ell(X^i_{k+1}, Y_{k+1})}{q_k(X^i_k, X^i_{k+1})} \; .$$

Sequential Smoothing through IS, or SIS (Handschin & Mayne, 1969-1970)

1. Propose n independent particle trajectories $\{X^i_{0:k+1}\}^{1\leq i\leq n}$ under a Markovian scheme such that

$$p(x_{0:k+1}) = \rho_{0:k+1}(x_{0:k+1}) = q_0(x_0)\prod_{l=1}^{k} q_l(x_l, x_{l+1}) \ .$$

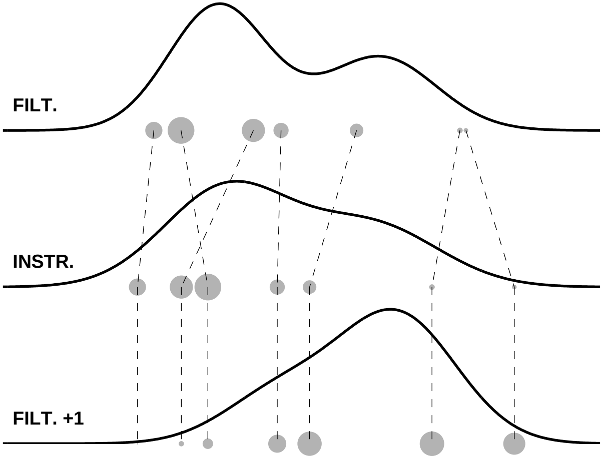
2. Compute importance weights sequentially:

$$\omega^i_{k+1} = \frac{\pi_{0:k+1|k+1}(X^i_{0:k+1})}{\rho_{0:k+1}(X^i_{0:k+1})} = \omega^i_k \times \frac{t(X^i_k, X^i_{k+1})\ell(X^i_{k+1}, Y_{k+1})}{q_k(X^i_k, X^i_{k+1})} \ .$$

Then,

$$\sum_{i=1}^{n} \frac{\omega^i_{k+1}}{\sum_{j=1}^{n}\omega^j_{k+1}} f(X^i_{0:k+1})$$

is an estimate of $\mathrm{E}\left[f(X_{0:k+1})\middle| Y_{0:k+1}\right]$.



One step of the SIS algorithm with just seven particles.

Weight Degeneracy

Empirically, the SIS approach always fail when the time-horizon k is more than a few tens; the IS weights $\omega_k^{1:n}$ usually become very unbalanced with a few weights dominating all the other

Weight Degeneracy

Empirically, the SIS approach always fail when the time-horizon k is more than a few tens; the IS weights $\omega_k^{1:n}$ usually become very unbalanced with a few weights dominating all the other

To understand why it is the case, consider the (silly) model where

$$\begin{cases} t(x, x') = t(x') = t_0(x') \ , & \text{(Independent states)} \\ \ell(x, y) = \ell(y) \ , & \text{(Non-informative observations)} \end{cases}$$

and the instrumental kernel is such that $q_l(x, x') = q_0(x') = q(x')$

Weight Degeneracy

Empirically, the SIS approach always fail when the time-horizon k is more than a few tens; the IS weights $\omega_k^{1:n}$ usually become very unbalanced with a few weights dominating all the other

To understand why it is the case, consider the (silly) model where

$$\begin{cases} t(x, x') = t(x') = t_0(x') \ , & \text{(Independent states)} \\ \ell(x, y) = \ell(y) \ , & \text{(Non-informative observations)} \end{cases}$$

and the instrumental kernel is such that $q_l(x, x') = q_0(x') = q(x')$

Then

$$\omega_{k+1}^i = \omega_k^i \times \frac{t(X_{k+1}^i)}{q(X_{k+1}^i)}$$

Weight Degeneracy (Contd.)

For a function of interest f that only depends on the last coordinate x_k of the trajectory $x_{0:k}$, the asymptotic variance of the SIS approximation to $\pi_{k|k}(f) = \mathrm{E}_{\pi_{k|k}}[f(X)]$ is given by

$$v_k(f) =$$

$$\int \cdots \int \left(\prod_{l=0}^{k} \frac{t}{q}(x_l) \right)^2 \left(f(x_k) - \pi_{k|k}(f) \right)^2 \prod_{l=0}^{k} q(x_l)\, dx_0 \ldots dx_k$$

$$= \Bigg(\underbrace{\int \frac{t}{q}(x) t(x) dx}_{>1} \Bigg)^k \int \frac{t}{q}(x) \left(f(x) - \pi_{k|k}(f) \right)^2 t(x) dx \, .$$

In practise, this situation can usually be detected by monitoring the *effective sample size* or *entropy* criterions, which become abnormally small.

Summary

- Sequential Importance Sampling (SIS) is based on simulating independent Markovian trajectories.
- SIS is bound to degenerate in the long-term (depends on everything, including n, but typically between 10 to 100 observations).

In IS, it is indeed possible to reset the weights to a constant value at the price of a, usually moderate, increase in variance.

Sampling Importance Resampling (Rubin, 1987)

Replace $\{X^{1:n}, W^{1:n}\}$ by $\{\tilde{X}^{1:\tilde{N}}, \tilde{W}^{1:\tilde{N}}\}$ such that the discrepancy between the resampled weights $\{\tilde{W}^{1:\tilde{N}}\}$ is reduced and $\sum_{i=1}^{\tilde{N}} \tilde{W}^i \delta_{\tilde{X}^i}$ is a good approximation to $\sum_{i=1}^{n} W^i \delta_{X^i}$.

In IS, it is indeed possible to reset the weights to a constant value at the price of a, usually moderate, increase in variance.

Sampling Importance Resampling (Rubin, 1987)

Replace $\{X^{1:n}, W^{1:n}\}$ by $\{\tilde{X}^{1:\tilde{N}}, \tilde{W}^{1:\tilde{N}}\}$ such that the discrepancy between the resampled weights $\{\tilde{W}^{1:\tilde{N}}\}$ is reduced and $\sum_{i=1}^{\tilde{N}} \tilde{W}^i \delta_{\tilde{X}^i}$ is a good approximation to $\sum_{i=1}^{n} W^i \delta_{X^i}$.

In general the resampling is random and subject to the constraints

$$\begin{cases} \tilde{N} = n \,, \\ \tilde{W}^i = 1/\tilde{N} \,, \\ \mathrm{E}\left[\sum_{i=1}^{\tilde{N}} \mathbb{1}\{\tilde{X}^i = X^j\} \middle| X^{1:n}, W^{1:n} \right] = \tilde{N} W^j \quad (1 \leq j \leq n). \end{cases}$$

The last condition is often referred to as *unbiasedness* or *proper weighting*.

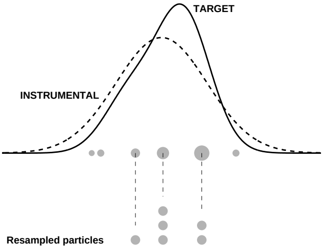
Multinomial Resampling

1. Draw, conditionally independently given $\{X^{1:n}, W^{1:n}\}$, n discrete random variables $(J^1, \ldots, J^n)$ taking their values in the set $\{1, \ldots, n\}$ with probabilities $(W^1, \ldots, W^n)$.
2. Set, for $i = 1, \ldots, n$, $\tilde{X}^i = X^{J^i}$ and $\tilde{W}^i = 1/n$.

Multinomial Resampling

1. Draw, conditionally independently given $\{X^{1:n}, W^{1:n}\}$, n discrete random variables $(J^1, \ldots, J^n)$ taking their values in the set $\{1, \ldots, n\}$ with probabilities $(W^1, \ldots, W^n)$.
2. Set, for $i = 1, \ldots, n$, $\tilde{X}^i = X^{J^i}$ and $\tilde{W}^i = 1/n$.

Let $C^i = \sum_{j=1}^n \mathbb{1}\{\tilde{X}^j = X^i\}$ $(i = 1, \ldots, n)$ denote the number of times each particle is duplicated in the resampling process. The counts $(C^1, \ldots, C^n)$ follow a multinomial distribution with parameters n, $(W^1, \ldots, W^n)$, conditionally to $\{X^{1:n}, W^{1:n}\}$.



Some Results on SIR

1. $\tilde{X}^i \xrightarrow{\mathcal{D}} \pi$ as $n \to \infty$ (some extensions of this result)
2. $\frac{1}{n}\sum_{i=1}^{n} f(\tilde{X}^i)$ is an asymptotically normal estimator of $\pi(f)$ (assuming $\mathrm{E}_\pi[\frac{\pi}{q}(X)(1+f^2(X))+f^2(X)] < \infty$) with asymptotic variance given by

$$\tilde{v}_q(f) = \underbrace{\mathrm{E}_\pi\left[\frac{\pi}{q}(X)\left(f(X)-\pi(f)\right)^2\right]}_{v_q(f)} + \underbrace{\mathrm{E}_\pi\left[\left(f(X)-\pi(f)\right)^2\right]}_{\mathrm{Var}_\pi[f(X)]}$$

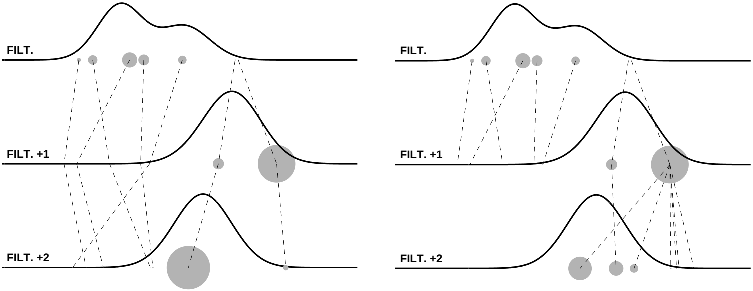
If n is sufficiently large, the cost of resampling is very moderate in situation that are challenging for IS, i.e., when $v_q(f) \gg \mathrm{Var}_\pi[f(X)]$.

The Simplest Functional Algorithm (Gordon et al., 1993)

Regular resampling is added to avoid weight degeneracy and to guarantee the long-term ($k \to \infty$) stability of the particle filter.

The Bootstrap filter

1. Given $\tilde{X}_k^{1:n}$, propose new positions X_{k+1}^i independently under the prior dynamic $t(\tilde{X}_k^i, \cdot)$, for $i = 1, \ldots, n$;
2. Compute the weights $\omega_{k+1}^i = \ell(X_{k+1}^i, Y_{k+1})$, for $i = 1, \ldots, n$ and normalize them ($W_{k+1}^i = \omega_{k+1}^i / \sum_{j=1}^n \omega_{k+1}^j$);
3. Resample to obtain $\tilde{X}_{k+1}^{1:n}$, e.g., by drawing independent indices J_{k+1}^i such that $\mathrm{P}\left(J_{k+1}^i = j \middle| W_{k+1}^{1:n}\right) = W_{k+1}^j$ and setting $\tilde{X}_{k+1}^i = X_{k+1}^{J_{k+1}^i}$ (Multinomial Resampling).



SIS (left) and SISR (right).

Marginal and Trajectory-Wise Approximations

SMC is expected to approximate the filtering pdfs in the sense that

$$\sum_{i=1}^{n} W_k^i f(X_k^i) \longrightarrow \int f(x)\pi_{k|k}(x)dx \ ,$$

as n increases, for abitrary functions f.

But recalling our original SIS interpretation, one should also have

$$\sum_{i=1}^{n} W_k^i f(X_{0:k}^i) \longrightarrow \int \cdots \int f(x_{0:k})\pi_{0:k|k}(x_{0:k})dx_0 \ldots dx_k \ .$$

Marginal and Trajectory-Wise Approximations

SMC is expected to approximate the filtering pdfs in the sense that

$$\sum_{i=1}^{n} W_k^i f(X_k^i) \longrightarrow \int f(x)\pi_{k|k}(x)dx \ ,$$

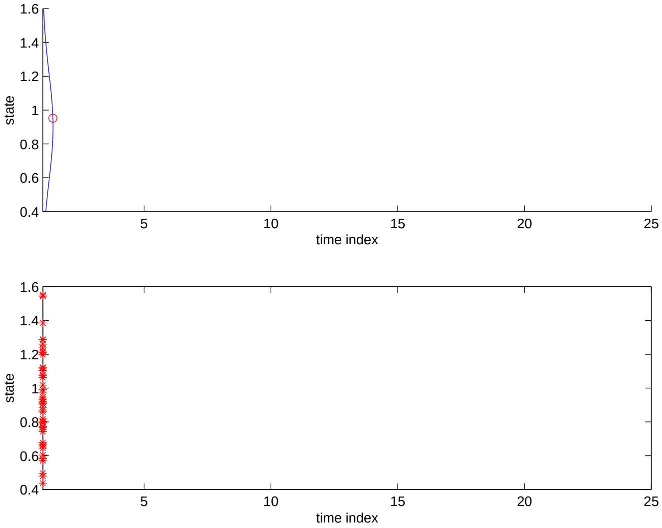
as n increases, for abitrary functions f.

But recalling our original SIS interpretation, one should also have

$$\sum_{i=1}^{n} W_k^i f(X_{0:k}^i) \longrightarrow \int \cdots \int f(x_{0:k})\pi_{0:k|k}(x_{0:k})dx_0 \dots dx_k \ .$$

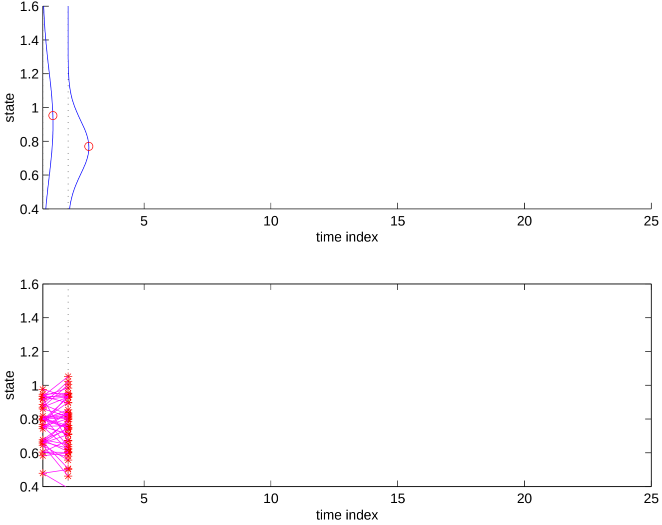
1. In what sense is this true? [Several: Consistency, central limit theorem, L^p bounds, convergence in distribution of subpopulations ("propagation of chaos")]
2. What is the influence of n? [Easy: $1\sqrt{n}$]
3. What is the influence of k? [Harder: depends on forgetting properties of the model and whether one considers marginal or trajectory-wise approximations]

The Particle Paths



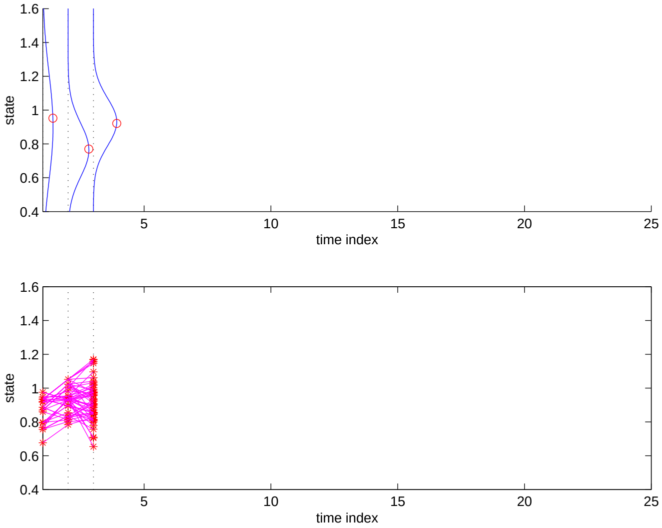
Predictive densities and evolution of the particle ancestry tree

The Particle Paths



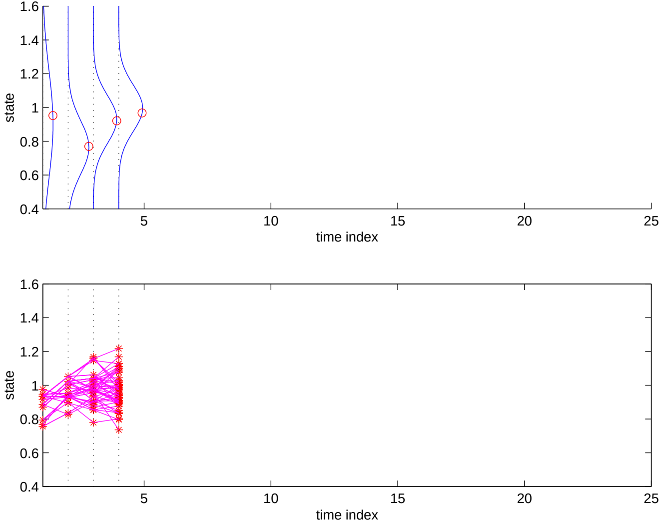
Predictive densities and evolution of the particle ancestry tree

The Particle Paths



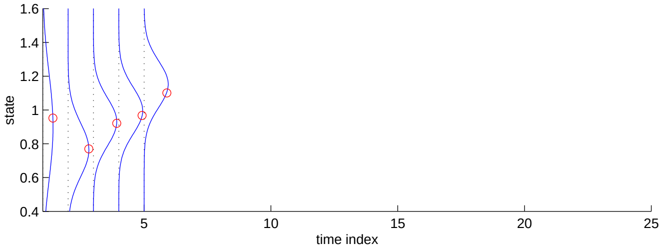
Predictive densities and evolution of the particle ancestry tree

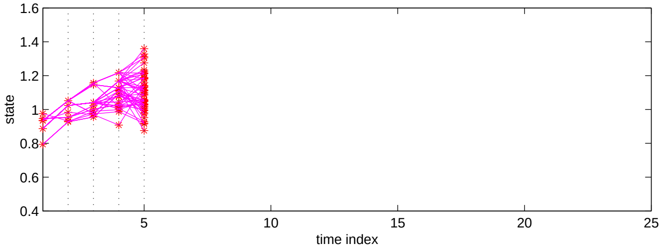
The Particle Paths



Predictive densities and evolution of the particle ancestry tree

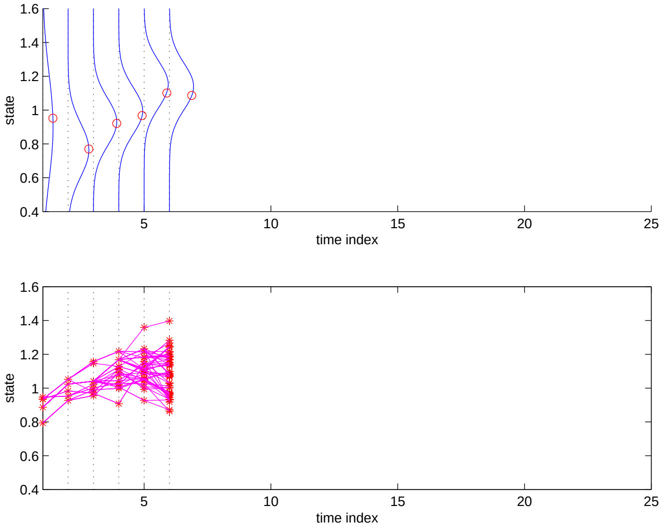
The Particle Paths





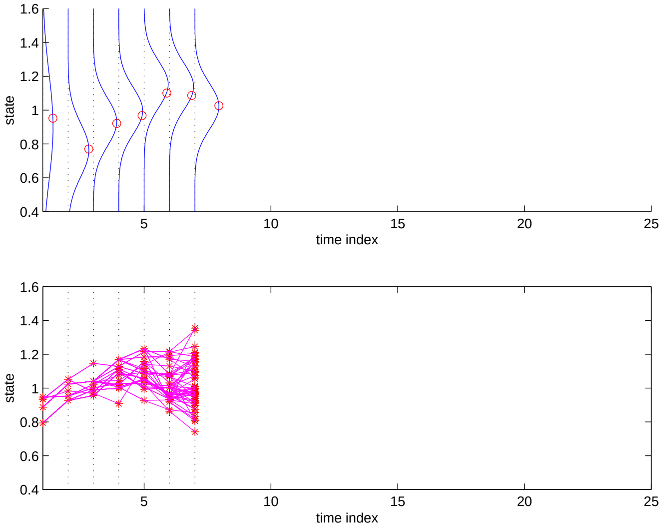
Predictive densities and evolution of the particle ancestry tree

The Particle Paths



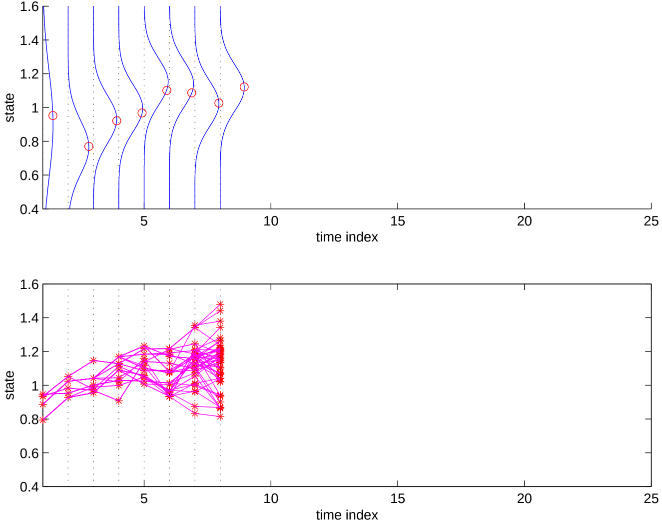
Predictive densities and evolution of the particle ancestry tree

The Particle Paths



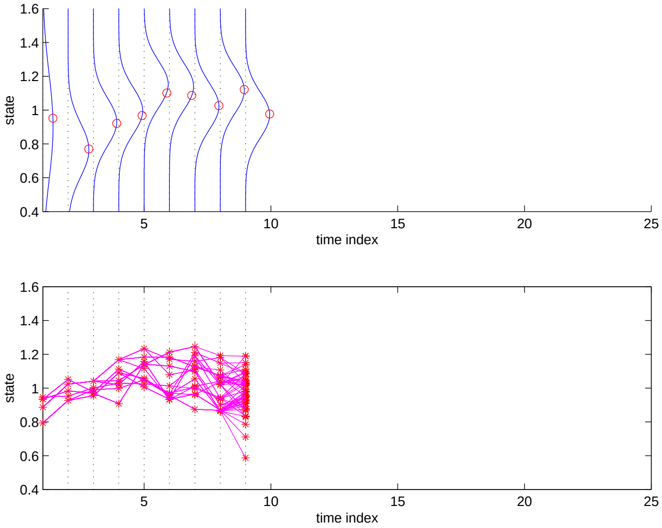
Predictive densities and evolution of the particle ancestry tree

The Particle Paths



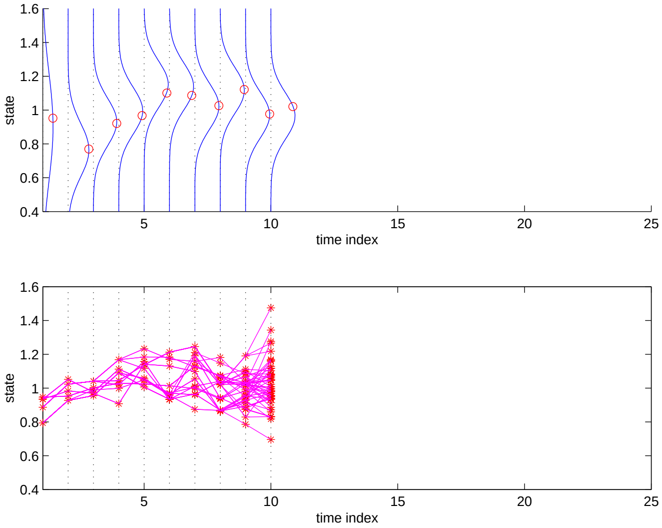
Predictive densities and evolution of the particle ancestry tree

The Particle Paths



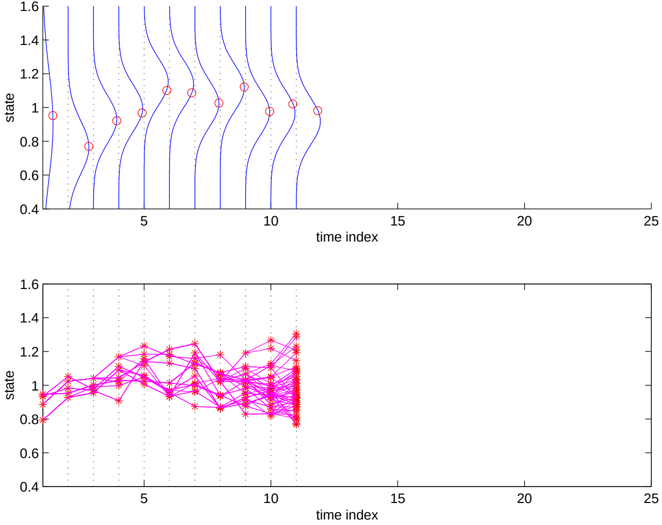
Predictive densities and evolution of the particle ancestry tree

The Particle Paths



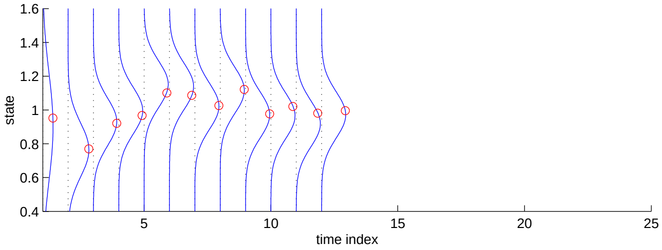
Predictive densities and evolution of the particle ancestry tree

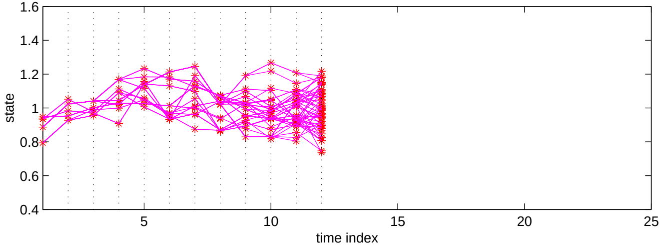
The Particle Paths



Predictive densities and evolution of the particle ancestry tree

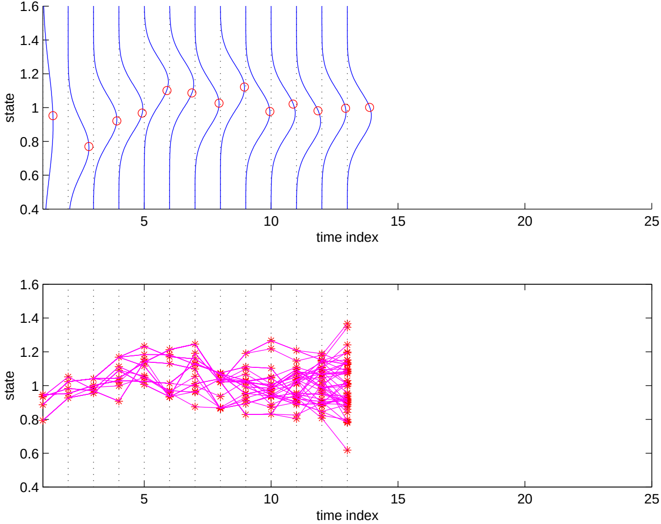
The Particle Paths





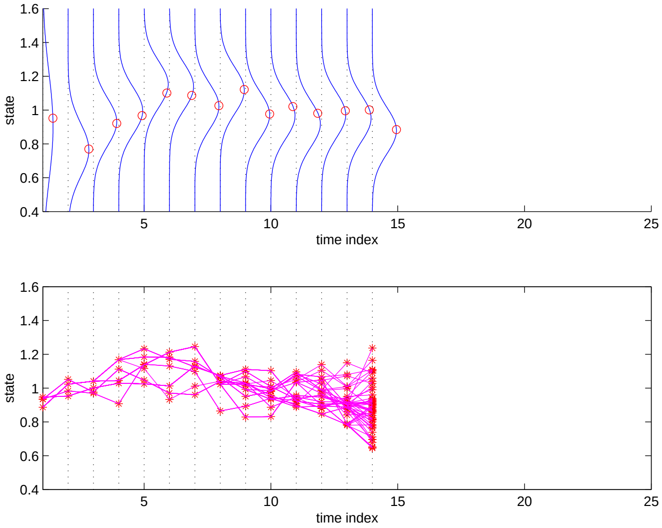
Predictive densities and evolution of the particle ancestry tree

The Particle Paths



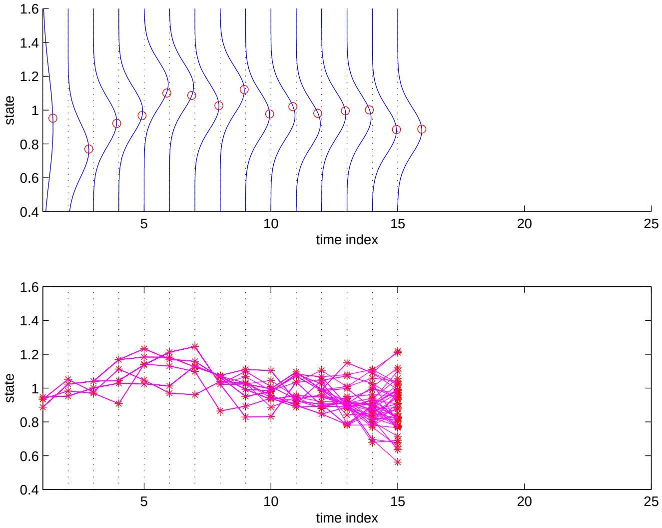
Predictive densities and evolution of the particle ancestry tree

The Particle Paths



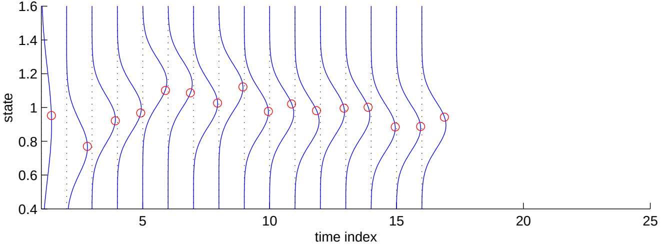
Predictive densities and evolution of the particle ancestry tree

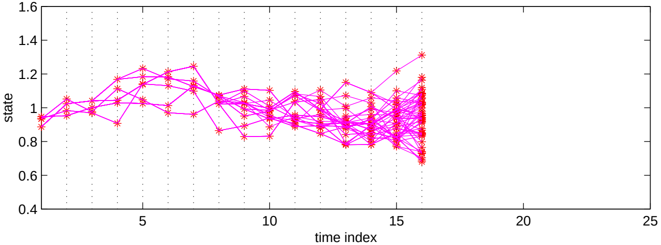
The Particle Paths



Predictive densities and evolution of the particle ancestry tree

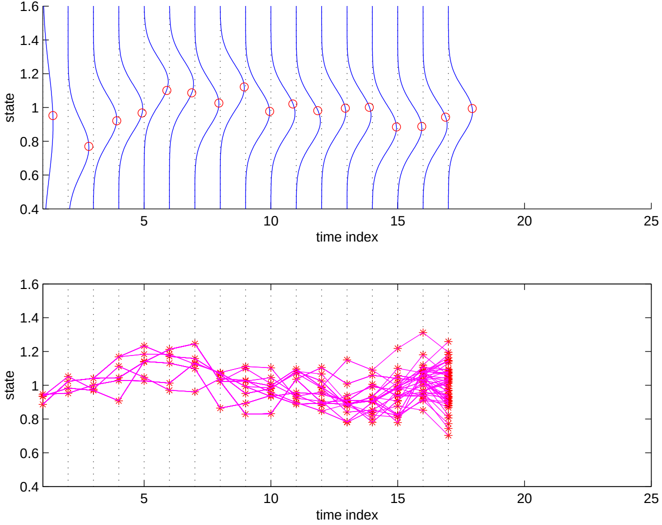
The Particle Paths





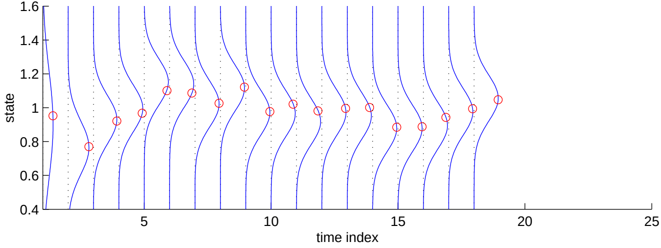
Predictive densities and evolution of the particle ancestry tree

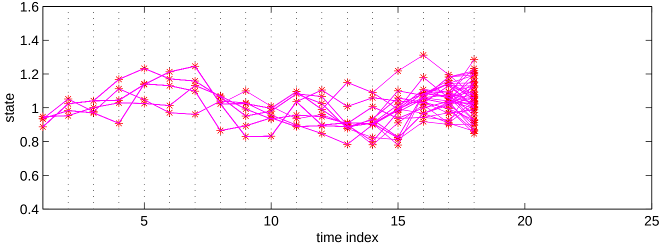
The Particle Paths



Predictive densities and evolution of the particle ancestry tree

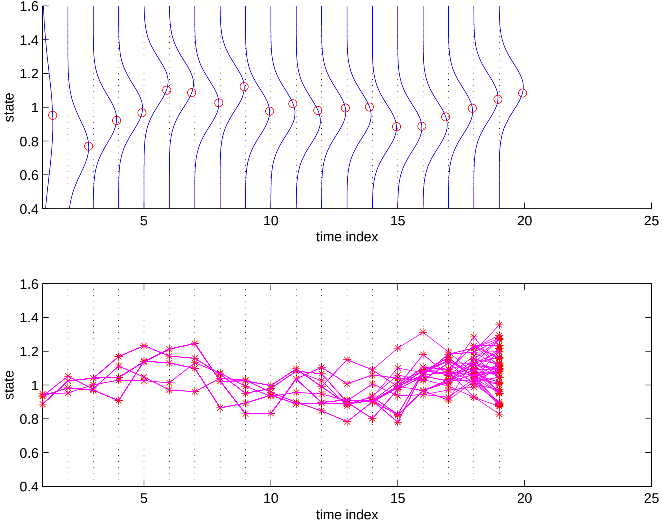
The Particle Paths





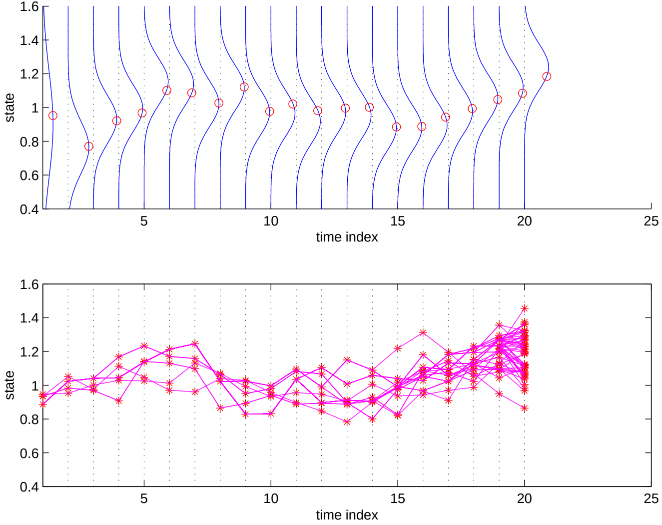
Predictive densities and evolution of the particle ancestry tree

The Particle Paths



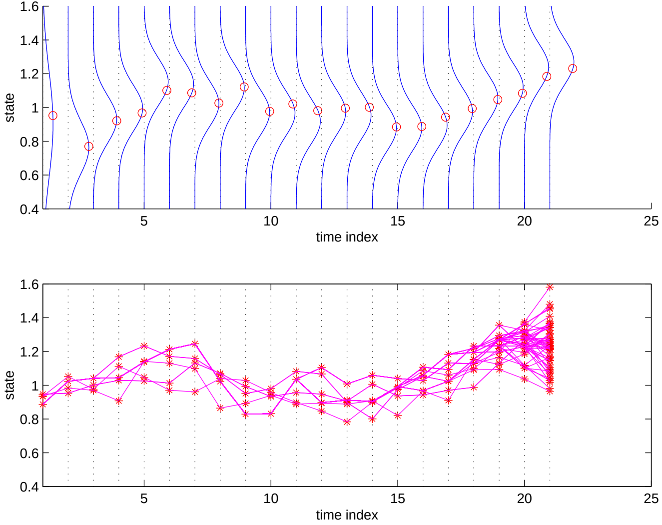
Predictive densities and evolution of the particle ancestry tree

The Particle Paths



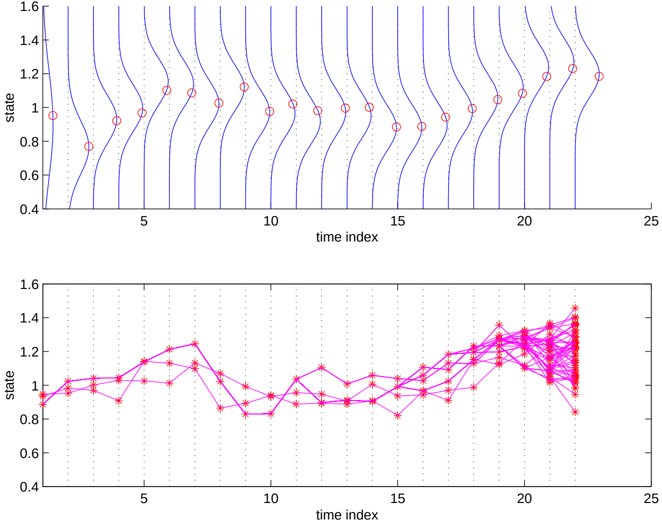
Predictive densities and evolution of the particle ancestry tree

The Particle Paths



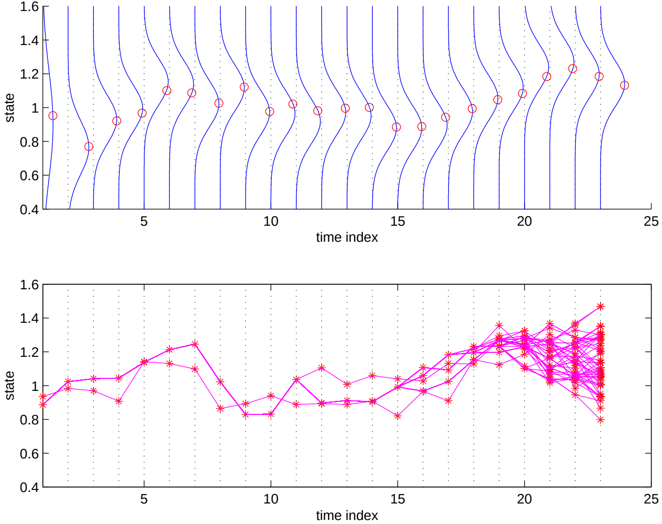
Predictive densities and evolution of the particle ancestry tree

The Particle Paths



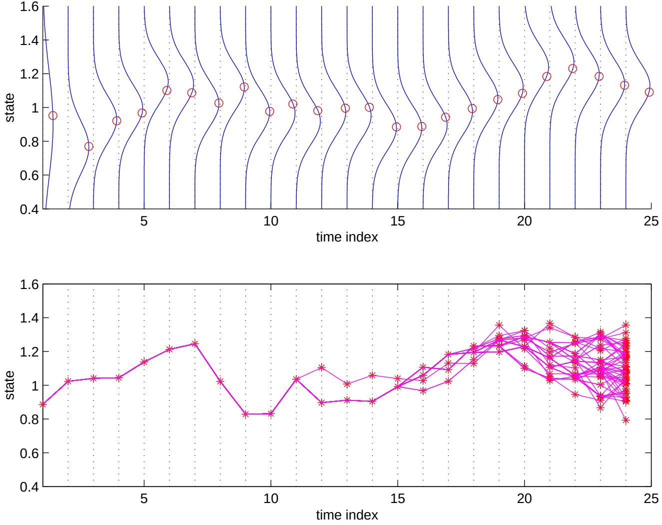
Predictive densities and evolution of the particle ancestry tree

The Particle Paths



Predictive densities and evolution of the particle ancestry tree

The Particle Paths



Predictive densities and evolution of the particle ancestry tree

Summary

- With resampling, SISR can achieve long-term stability.
- The increase in variance due to resampling is moderate, especially when resampling is applied only when needed.
- The method is still sensitive to outliers, model misspecification, etc., which may necessitate the use of more elaborate strategies (clever choices of the instrumental kernel, adaptive strategies, etc.)
- Accurate smoothing approximations require more elaborate techniques

- Doucet, A., De Freitas, N. and Gordon, N. (eds.) (2001), *Sequential Monte Carlo Methods in Practice*, Springer.
- Ristic, B., Arulampalam, M. and Gordon, A. (2004), *Beyond Kalman Filters: Particle Filters for Target Tracking*, Artech House.
- Cappé, 0., Moulines, E. and Rydén, T. (2005), *Inference in Hidden Markov Models*, Springer.
- Doucet, A., Godsill, S. and Andrieu, C. (2000), On sequential Monte-Carlo sampling methods for Bayesian filtering, Stat. Comput., **10**:197–208.
- Arulampalam, M., Maskell, S., Gordon, N. and Clapp, T. (2002), A tutorial on particle filters for on line non-linear/non-Gaussian Bayesian tracking, IEEE Trans. Signal Process., **50**, 241–254.
- Cappé, O., Godsill, S. J. and Moulines, E. (2007), An overview of existing methods and recent advances in sequential Monte Carlo, IEEE Proc., **95**:899–924.