Long-term stability of sequential Monte Carlo methods under verifiable conditions

Jimmy Olsson
Centre for Mathematical Sciences
Lund University

University of Copenhagen
April 11, 2012
Plan of the talk

1. Hidden Markov models

2. Particle filters
   - The bootstrap particle filter
   - Uniform convergence of particle filters

3. Some novel results
   - Aims
   - Assumptions
   - Main results

4. Elements of proof
1 Hidden Markov models

2 Particle filters
   - The bootstrap particle filter
   - Uniform convergence of particle filters

3 Some novel results
   - Aims
   - Assumptions
   - Main results

4 Elements of proof
Hidden Markov models (HMMs)

An HMM comprises

- an unobservable Markov chain $X \triangleq (X_n)_{n \geq 0}$ on $(\mathcal{X}, \mathcal{X})$ with transition kernel $Q$ and initial distribution $\chi$, i.e. $X_0 \sim \chi$ and

$$
\mathbb{P}(X_{n+1} \in A | \mathcal{F}_n) = \mathbb{P}(X_{n+1} \in A | X_n) = Q(X_n, A), \quad A \in \mathcal{X}.
$$

$\chi$
An HMM comprises

- an unobservable Markov chain $X \triangleq (X_n)_{n \geq 0}$ on $(\mathcal{X}, \mathcal{X}')$ with transition kernel $Q$ and initial distribution $\chi$, i.e. $X_0 \sim \chi$ and

$$
P(X_{n+1} \in A | \mathcal{F}_n) = P(X_{n+1} \in A | X_n) = Q(X_n, A), \quad A \in \mathcal{X}.
$$

- observations $(Y_n)_{n \geq 0}$ in $(\mathcal{Y}, \mathcal{Y})$ being conditionally independent given $X$ such that

$$
P(Y_n \in B | X) = P(Y_n \in B | X_n) = G(X_n, B), \quad B \in \mathcal{Y}.
$$

We assume that $G$ has a transition density $g$, i.e.

$$G(x, B) = \int_B g(x, y) \nu(dy).$$
Graphical representation:

\[ Y_{n-1} \rightarrow X_{n-1} \rightarrow X_n \rightarrow X_{n+1} \rightarrow Y_n \rightarrow Y_{n+1} \]

(Observations)

(Markov chain)

\[ Y_n | X_n \sim G(X_n, \cdot) \]
\[ X_{n+1} | X_n \sim Q(X_n, \cdot) \]
\[ X_0 \sim \chi \]
Any kind of estimation (filtering, parameter inference) in HMMs typically involves the computation of the predictor distributions

\[ \phi_X \langle Y_{0}^{n-1} \rangle (A) = \mathbb{P}(X_n \in A | Y_{0}^{n-1}) \]

expressed recursively through the predictor recursion

\[ \phi_X \langle Y_{0}^{n} \rangle (A) = \frac{\int g(x_n, Y_n) Q(x_n, A) \phi_X \langle Y_{0}^{n-1} \rangle (dx_n)}{\int g(x_n, Y_n) \phi_X \langle Y_{0}^{n-1} \rangle (dx_n)}, \]
Any kind of estimation (filtering, parameter inference) in HMMs typically involves the computation of the predictor distributions

\[ \phi \chi \langle Y_0^{n-1} \rangle (A) = \mathbb{P}(X_n \in A | Y_0^{n-1}) \]

expressed recursively through the predictor recursion

\[ \phi \chi \langle Y_0^n \rangle (A) = \frac{\int g(x_n, Y_n) Q(x_n, A) \phi \chi \langle Y_0^{n-1} \rangle (dx_n)}{\int g(x_n, Y_n) \phi \chi \langle Y_0^{n-1} \rangle (dx_n)}, \]

defining a measure-valued mapping \( \Phi \langle Y_n \rangle \) such that

\[ \phi \chi \langle Y_0^n \rangle = \Phi \langle Y_n \rangle (\phi \chi \langle Y_0^{n-1} \rangle). \]
Predictor distributions

Any kind of estimation (filtering, parameter inference) in HMMs typically involves the computation of the predictor distributions

\[ \phi_{\chi}\langle Y_{0}^{n-1}\rangle(A) = \mathbb{P}(X_n \in A|Y_{0}^{n-1}) \]

expressed recursively through the predictor recursion

\[ \phi_{\chi}\langle Y_{0}^{n}\rangle(A) = \frac{\int g(x_n, Y_n)Q(x_n, A) \phi_{\chi}\langle Y_{0}^{n-1}\rangle(dx_n)}{\int g(x_n, Y_n) \phi_{\chi}\langle Y_{0}^{n-1}\rangle(dx_n)}, \]

defining a measure-valued mapping \( \Phi\langle Y_n \rangle \) such that

\[ \phi_{\chi}\langle Y_{0}^{n}\rangle = \Phi\langle Y_n \rangle(\phi_{\chi}\langle Y_{0}^{n-1}\rangle). \]

However, the recursion lacks closed-form solution in general.
We are here

1. Hidden Markov models

2. Particle filters
   - The bootstrap particle filter
   - Uniform convergence of particle filters

3. Some novel results
   - Aims
   - Assumptions
   - Main results

4. Elements of proof
Particle filters in a nutshell

A particle filter generates and updates recursively a set of particles $(\xi_n^i)_{i=1}^N$ such that for all bounded measurable functions $h$,

$$\frac{1}{N} \sum_{i=1}^{N} h(\xi_n^i) \overset{\text{not.}}{=} \phi_X^N \langle Y_0^{n-1} \rangle h \approx \int h(x) \phi_X^N \langle Y_0^{n-1} \rangle (dx) \overset{\text{not.}}{=} \phi_X \langle Y_0^{n-1} \rangle h.$$

Comprises in general two main operations, namely

1. a mutation step that disseminates the particles randomly in $X$
   and

2. a selection step that duplicates/eliminates particles with high/low posterior probability.
Hidden Markov models

Particle filters

Some novel results

Elements of proof

1. Hidden Markov models

2. Particle filters
   - The bootstrap particle filter
   - Uniform convergence of particle filters

3. Some novel results
   - Aims
   - Assumptions
   - Main results

4. Elements of proof
The bootstrap particle filter

Let \( (\xi^i_n)_{i=1}^N \) be a particle approximation of \( \phi \langle Y_0^{n-1} \rangle \). To update the particles as \( Y_n \) appears,

- **(weighting)** compute the importance weights

\[
\omega_n^i = g(\xi_n^i, Y_n)
\]

- **(selection + mutation)** and draw \( (\xi^i_{n+1})_{i=1}^N \) from the mixture

\[
\sum_{i=1}^N \frac{\omega_n^i}{\sum_{\ell=1}^N \omega_{n,\ell}} Q(\xi_n^i, \cdot).
\]

In fact, this corresponds to sampling

\[
(\xi^i_{n+1})_{i=1}^N \overset{\text{i.i.d.}}{\sim} \Phi \langle Y_n \rangle (\phi^N \langle Y_0^{n-1} \rangle).
\]
Film time! 😊
Hidden Markov models
Particle filters
Some novel results
Elements of proof

The bootstrap particle filter
Uniform convergence of particle filters

J. Olsson

Long-term stability of SMC methods
Hidden Markov models
Particle filters
Some novel results
Elements of proof

The bootstrap particle filter
Uniform convergence of particle filters

→ prédiction
correction
sélection

J. Olsson
Long-term stability of SMC methods
Hidden Markov models
Particle filters
Some novel results
Elements of proof

The bootstrap particle filter
Uniform convergence of particle filters

J. Olsson
Long-term stability of SMC methods
Hidden Markov models
Particle filters
Some novel results
Elements of proof

The bootstrap particle filter
Uniform convergence of particle filters

→ prédiction
correction
sélection

J. Olsson
Long-term stability of SMC methods
Hidden Markov models
Particle filters
Some novel results
Elements of proof

The bootstrap particle filter
Uniform convergence of particle filters

J. Olsson
Long-term stability of SMC methods
Hidden Markov models
Particle filters
Some novel results
Elements of proof

The bootstrap particle filter
Uniform convergence of particle filters

J. Olsson
Long-term stability of SMC methods
Hidden Markov models
Particle filters
Some novel results
Elements of proof
The bootstrap particle filter
Uniform convergence of particle filters

J. Olsson
Long-term stability of SMC methods
Hidden Markov models
Particle filters
Some novel results
Elements of proof

The bootstrap particle filter
Uniform convergence of particle filters

J. Olsson

Long-term stability of SMC methods
Hidden Markov models
Particle filters
Some novel results
Elements of proof
The bootstrap particle filter
Uniform convergence of particle filters
Long-term stability of SMC methods
Hidden Markov models
Particle filters
Some novel results
Elements of proof

The bootstrap particle filter
Uniform convergence of particle filters

J. Olsson
Long-term stability of SMC methods
Hidden Markov models
Particle filters
Some novel results
Elements of proof

The bootstrap particle filter
Uniform convergence of particle filters

J. Olsson

Long-term stability of SMC methods
Hidden Markov models
Particle filters
Some novel results
Elements of proof

The bootstrap particle filter
Uniform convergence of particle filters

J. Olsson
Long-term stability of SMC methods
Hidden Markov models
Particle filters
Some novel results
Elements of proof

The bootstrap particle filter
Uniform convergence of particle filters

→ prédiction
 correction
 sélection

J. Olsson
Long-term stability of SMC methods
Hidden Markov models
Particle filters
Some novel results
Elements of proof

The bootstrap particle filter
Uniform convergence of particle filters

J. Olsson
Long-term stability of SMC methods
Hidden Markov models
Particle filters
Some novel results
Elements of proof

The bootstrap particle filter
Uniform convergence of particle filters

J. Olsson
Long-term stability of SMC methods
Hidden Markov models
Particle filters
Some novel results
Elements of proof

The bootstrap particle filter
Uniform convergence of particle filters

prédiction
correction
→ sélection

J. Olsson

Long-term stability of SMC methods
Hidden Markov models
Particle filters
Some novel results
Elements of proof

The bootstrap particle filter
Uniform convergence of particle filters

J. Olsson
Long-term stability of SMC methods
Hidden Markov models
Particle filters
Some novel results
Elements of proof

The bootstrap particle filter
Uniform convergence of particle filters

Long-term stability of SMC methods
1 Hidden Markov models

2 Particle filters
   - The bootstrap particle filter
   - Uniform convergence of particle filters

3 Some novel results
   - Aims
   - Assumptions
   - Main results

4 Elements of proof
There is number of convergence results concerning e.g. $L^p$ error and weak convergence (see e.g. Del Moral, 2004, Bain and Crisan, 2009).

Most of these results establish the convergence, as $N$ tends to infinity, for fixed, finite time steps $n$. 
Convergence of particle filters

- There is number of convergence results concerning e.g. $L^p$ error and weak convergence (see e.g. Del Moral, 2004, Bain and Crisan, 2009).
- Most of these results establish the convergence, as $N$ tends to infinity, for fixed, finite time steps $n$.
- However, for infinite time horizons, i.e. when $n$ tends to infinity, things are not that obvious, since we may expect an accumulation of error.
- **Crucial question:** Under which conditions is the error propagated through the algorithm controlled as $n$ increases?
Example: Linear Gaussian HMM

Comparison of standard sequential IS (○) and the bootstrap filter (⋆) with exact values (⋆) provided by the Kalman filter:

Filtered means for Kalman filter, SIS, and SISR
Error decomposition

In the decomposition

$$\phi^N_Y \langle Y^n \rangle - \phi^N_X \langle Y^n \rangle = \underbrace{\phi^N_X \langle Y^n_0 \rangle - \Phi \langle Y_n \rangle (\phi^N_X \langle Y^{n-1}_0 \rangle)}_{\text{local error}} + \underbrace{\Phi \langle Y_n \rangle (\phi^N_X \langle Y^{n-1}_0 \rangle) - \Phi \langle Y_n \rangle (\phi^N_X \langle Y^{n-1}_0 \rangle)}_{\text{propagation error}},$$

- the **local error** stems from replacing $$\Phi \langle Y_n \rangle (\phi^N_X \langle Y^{n-1}_0 \rangle)$$ by the empirical estimate $$\phi^N_X \langle Y^n_0 \rangle$$.
- the **propagation error** originates from the discrepancy between $$\phi^N_X \langle Y^{n-1}_0 \rangle$$ and $$\phi^N_X \langle Y^{n-1}_0 \rangle$$, but mapped through the predictor recursion $$\Phi$$. 
Here

- the **local error** is easy handled, since the particles \((\xi_i^n)_{i=1}^N\) are conditionally independent with distribution \(\Phi(Y_n)\langle \phi_X \langle Y_{n-1} \rangle \rangle\).

- The **Marcinkiewicz-Zygmund inequality** provides a bound on the \(L^p\) error of form \(c/\sqrt{N}\).
- **Hoeffding’s inequality** provides an exponential deviation inequality.
Here

- the **local error** is easy handled, since the particles \( (\xi_n^i)_{i=1}^N \) are conditionally independent with distribution \( \Phi(Y_n)(\phi^N_Y(Y_0^{n-1})) \).

- The **Marcinkiewicz-Zygmund inequality** provides a bound on the \( L^p \) error of form \( c/\sqrt{N} \).

- **Hoeffding’s inequality** provides an exponential deviation inequality.

- the **propagation error** can be controlled if the mapping \( \Phi \) is in some sense contracting and thus downweights the discrepancy between \( \phi^N_Y(Y_0^{n-1}) \) and \( \phi_Y(Y_0^{n-1}) \).
Error decomposition (cont.)

- Here the **local error** is easy handled, since the particles $(\xi_n^i)_{i=1}^N$ are conditionally independent with distribution $\Phi\langle Y_n \rangle (\phi_N^N \langle Y_0^{n-1} \rangle)$.

- The **Marcinkiewicz-Zygmund inequality** provides a bound on the $L^p$ error of form $c/\sqrt{N}$.

- **Hoeffding’s inequality** provides an exponential deviation inequality.

- The **propagation error** can be controlled if the mapping $\Phi$ is in some sense contracting and thus downweights the discrepancy between $\phi_N^N \langle Y_0^{n-1} \rangle$ and $\phi_\chi \langle Y_0^{n-1} \rangle$.

- The **forgetting** of the initial distribution in the predictor recursion becomes crucial.
Stability of the predictor recursion

Assume that there exist constants $C < \infty$ and $\rho \in (0, 1)$ such that

$$||\Phi\langle Y^n_m \rangle(\chi) - \Phi\langle Y^n_m \rangle(\chi')|| \leq C\rho^{n-m+1}||\chi - \chi'||$$

for all probability measures $\chi$ and $\chi'$, where $|| \cdot ||$ is some suitable norm on the space of probability measures and

$$\Phi\langle Y^n_m \rangle \triangleq \Phi\langle Y^n \rangle \circ \Phi\langle Y_{n-1} \rangle \circ \cdots \circ \Phi\langle Y_m \rangle.$$
Assume that there exist constants $C < \infty$ and $\rho \in (0, 1)$ such that

$$\| \Phi\langle Y^n_m \rangle(\chi) - \Phi\langle Y^n_m \rangle(\chi') \| \leq C \rho^{n-m+1} \| \chi - \chi' \|$$

for all probability measures $\chi$ and $\chi'$, where $\| \cdot \|$ is some suitable norm on the space of probability measures and

$$\Phi\langle Y^n_m \rangle \triangleq \Phi\langle Y^n_n \rangle \circ \Phi\langle Y^n_{n-1} \rangle \circ \cdots \circ \Phi\langle Y^n_m \rangle.$$ 

Since $\Phi\langle Y^n_m \rangle(\chi) = \phi_\chi\langle Y^n_m \rangle$, this means that the predictor distribution 
forgets the initial distribution geometrically fast with a rate that is uniform w.r.t. initial distributions and observations.
Iterating the error decomposition $n$ times yields the telescoping sum

$$
\phi^N \langle Y^n \rangle - \phi \langle Y^n \rangle = \phi^N \langle Y^0 \rangle - \Phi \langle Y_n \rangle (\phi^N \langle Y^{n-1} \rangle)
$$

$$
+ \sum_{k=1}^{n-1} \left( \Phi \langle Y^k_{n+1} \rangle (\phi^N \langle Y^k \rangle) - \Phi \langle Y^k_{n+1} \rangle \circ \Phi \langle Y_{n+1} \rangle (\phi^N \langle Y^{k-1} \rangle) \right)
$$

$$
+ \Phi \langle Y^n_1 \rangle (\phi^N \langle Y_0 \rangle) - \Phi \langle Y_1^n \rangle (\phi \langle Y_0 \rangle).
$$
Iterating the error decomposition \( n \) times yields the telescoping sum

\[
\phi^N_X \langle Y^n_0 \rangle - \phi_X \langle Y^n_0 \rangle = \phi^N_X \langle Y^n_0 \rangle - \Phi \langle Y_n \rangle (\phi^N_X \langle Y^{n-1}_0 \rangle) \\
+ \sum_{k=1}^{n-1} \left( \Phi \langle Y^n_{k+1} \rangle (\phi^N_X \langle Y^k_0 \rangle) - \Phi \langle Y^n_{k+1} \rangle \circ \Phi \langle Y_k \rangle (\phi^N_X \langle Y^{k-1}_0 \rangle) \right) \\
+ \Phi \langle Y^n_1 \rangle (\phi^N_X \langle Y^0_0 \rangle) - \Phi \langle Y^n_1 \rangle (\phi_X \langle Y^0_0 \rangle).
\]

Each term of this sum can be viewed as a downweighting (by a factor \( C \rho^{n-k} \)) of the local sampling error between \( \phi^N_X \langle Y^k_0 \rangle \) and \( \Phi \langle Y_k \rangle (\phi_X \langle Y^{k-1}_0 \rangle) \) through the contraction of \( \Phi \langle Y^n_{k+1} \rangle \).
The decomposition gives

\[ L^p \text{ error of } \phi_N^0 \langle Y^n \rangle \leq \frac{c}{\sqrt{N}} + \sum_{k=0}^{n-1} \frac{Cc}{\sqrt{N}} \rho^{n-k} + \frac{Cc}{\sqrt{N}} \rho^n \]

\[ \leq \frac{C'}{\sqrt{N}(1 - \rho)} , \]

providing a bound that is uniform in \( n \) as well as the observations.
The decomposition gives

\[ L^p \text{ error of } \phi^N_N(\langle Y_0^n \rangle) \leq \frac{c}{\sqrt{N}} + \sum_{k=0}^{n-1} \frac{Cc}{\sqrt{N}} \rho^{n-k} + \frac{Cc}{\sqrt{N}} \rho^n \]

\[ \leq \frac{C'}{\sqrt{N}(1 - \rho)}, \]

providing a bound that is uniform in \( n \) as well as the observations.

Here it is essential that

- the constant \( c \) in the M-Z inequality is universal and that
- the contracting rate \( C\rho^{n-k} \) is uniform in the initial distributions and well as the observations.
The global Doeblin condition

The Markov chain $X$ satisfies a global Doeblin condition if there exist constants $0 < \epsilon^- < \epsilon^+$ and a probability measure $\lambda$ on $(X, \mathcal{X})$ such that for all $x \in X$ and $A \in \mathcal{X}$,

$$\epsilon^- \lambda(A) \leq Q(x, A) \leq \epsilon^+ \lambda(A).$$
The global Doeblin condition

The Markov chain $X$ satisfies a global Doeblin condition if there exist constants $0 < \epsilon^- < \epsilon^+$ and a probability measure $\lambda$ on $(X, \mathcal{X})$ such that for all $x \in X$ and $A \in \mathcal{X}$,

$$\epsilon^- \lambda(A) \leq Q(x, A) \leq \epsilon^+ \lambda(A).$$

This condition

1. implies that $X$ is uniformly geometrically ergodic.
2. implies that the predictor forgets the initial distribution geometrically fast with uniform rate $C \rho^n$, where

$$\rho = 1 - \frac{\epsilon^-}{\epsilon^+}.$$

3. is restrictive and points to applications where $X$ is compact, which is not the case in most applications of interest.
Hidden Markov models

Particle filters

Some novel results

Elements of proof

1. Hidden Markov models

2. Particle filters
   - The bootstrap particle filter
   - Uniform convergence of particle filters

3. Some novel results
   - Aims
   - Assumptions
   - Main results

4. Elements of proof
Hidden Markov models

Particle filters

Some novel results

Elements of proof

We are here

1 Hidden Markov models

2 Particle filters
   • The bootstrap particle filter
   • Uniform convergence of particle filters

3 Some novel results
   • Aims
   • Assumptions
   • Main results

4 Elements of proof
Our approach

The proof sketch above (inherited from Del Moral and Guionnet, 2001)

- required uniform forgetting w.r.t. the initial distributions. This will most presumably not be the case under more sensible assumptions (read: the non-compact case).
- provided a deterministic upper bound on the $L^p$ error. This is probably too much to ask for in the general case.
Our approach (cont.)

Our approach takes the following CLT (which can be established under weak assumptions) as starting point.

**Theorem (Del Moral and Guionnet, 1999)**

For all \( y_{0}^{n-1} \in Y^n \) and bounded measurable functions \( h \),

\[
\sqrt{N}(\phi_{\chi}^{N} \langle y_{0}^{n-1} \rangle h - \phi_{\chi} \langle y_{0}^{n-1} \rangle h) \Rightarrow_{N \to \infty} \mathcal{N}(0, \sigma^2_{\chi} \langle y_{0}^{n-1} \rangle(h)).
\]
Our approach (cont.)

Our approach takes the following CLT (which can be established under weak assumptions) as starting point.

**Theorem (Del Moral and Guionnet, 1999)**

For all $y_{0}^{n-1} \in \mathcal{Y}^{n}$ and bounded measurable functions $h$,

$$
\sqrt{N}(\phi_{\chi}^{N}\langle y_{0}^{n-1}\rangle h - \phi_{\chi}\langle y_{0}^{n-1}\rangle h) \Rightarrow_{N \to \infty} \mathcal{N}(0, \sigma_{\chi}^{2}\langle y_{0}^{n-1}\rangle(h)).
$$

Here the asymptotic variance $\sigma_{\chi}^{2}\langle y_{0}^{n-1}\rangle(h)$

- has a very complicated (recursive) form.
- can be uniformly bounded under the global Doeblin condition above (Del Moral and Guionnet, 2001).
Our approach (cont.)

Our approach takes the following CLT (which can be established under weak assumptions) as starting point.

\[
\sqrt{N} (\phi^N \langle y^{n-1}_0 \rangle h - \phi \langle y^{n-1}_0 \rangle h) \Rightarrow_{N \to \infty} N(0, \sigma^2 \langle y^{n-1}_0 \rangle (h)).
\]

Here the asymptotic variance \( \sigma^2 \langle y^{n-1}_0 \rangle (h) \)
- has a very complicated (recursive) form.
- can be uniformly bounded under the global Doeblin condition above (Del Moral and Guionnet, 2001).

As we will see, one may show that the sequence \( (\sigma^2 \langle h \rangle (Y^{n-1}_0))_{n \geq 0} \)
is tight under considerably weaker assumptions, implying in turn tightness of the asymptotic \( L^p \) error!
1. Hidden Markov models

2. Particle filters
   - The bootstrap particle filter
   - Uniform convergence of particle filters

3. Some novel results
   - Aims
   - Assumptions
   - Main results

4. Elements of proof
The local Doeblin condition

**Definition**

A set $C \in \mathcal{X}$ is a **local Doeblin** if there exist positive constants $0 < \epsilon^-_C < \epsilon^+_C$ and a probability measure $\lambda_C$ such that $\lambda_C(C) = 1$ and for all $A \in \mathcal{X}$ and $x \in C$,

$$\epsilon^-_C \lambda_C(A) \leq Q(x, A \cap C) \leq \epsilon^+_C \lambda_C(A).$$
The local Doeblin condition

**Definition**

A set $C \in \mathcal{X}$ is a **local Doeblin** if there exist positive constants $0 < \epsilon^-_C < \epsilon^+_C$ and a probability measure $\lambda_C$ such that $\lambda_C(C) = 1$ and for all $A \in \mathcal{X}$ and $x \in C$,

$$\epsilon^-_C \lambda_C(A) \leq Q(x, A \cap C) \leq \epsilon^+_C \lambda_C(A).$$

This condition

- is much weaker and covers many practical applications!
- is satisfied if $Q$ has a transition density $q$ such that

$$\frac{\sup_{(x,x') \in C^2} q(x, x')}{\inf_{(x,x') \in C^2} q(x, x')} < \infty,$$

which is the case if, for instance, $C$ is compact and $q$ is continuous.
Main assumptions

\((A1)\) The process \((Y_n)_{n \geq 0}\) is strictly stationary. In addition, there exists a set \(K \in \mathcal{Y}\) such that

(i) \(\mathbb{P}(Y_0 \in K) > 2/3\).

(ii) For all \(\eta > 0\) there exists a local Doeblin set \(C \in \mathcal{X}'\) such that for all \(y \in K\),

\[ \sup_{x \in C^c} g(x, y) \leq \eta \|g(\cdot, y)\|_\infty < \infty. \]

\((A2)\) (i) For all \((x, y) \in \mathcal{X} \times \mathcal{Y}, g(x, y) > 0.\)

(ii) \(\mathbb{E}(\ln^+ \|g(\cdot, Y_0)\|_\infty) < \infty.\)

\((A3)\) There exists a set \(D \in \mathcal{X}\) such that

(i) \(\inf_{x \in D} Q(x, D) > 0,\)

(ii) \(\mathbb{E}(\ln^- \inf_{x \in D} g(x, Y_0)) < \infty.\)
1. Hidden Markov models

2. Particle filters
   - The bootstrap particle filter
   - Uniform convergence of particle filters

3. Some novel results
   - Aims
   - Assumptions
   - Main results

4. Elements of proof

J. Olsson
Long-term stability of SMC methods
Main results

**Theorem**

Assume (A1)–(A3). Then for all bounded measurable functions \( h \), the sequence \( \sigma^2_X \langle Y_0^{n-1} \rangle(h) \) for \( n \geq 0 \) is tight.
Main results

**Theorem**

Assume (A1)–(A3). Then for all bounded measurable functions \( h \), the sequence \( (\sigma^2 \langle Y_0^{n-1} \rangle (h))_{n \geq 0} \) is tight.

**Theorem**

Under weak conditions it holds for all bounded measurable functions \( h \), \( p > 0 \), and initial distributions \( \chi \) such that \( \chi(D) > 0 \), \( \mathbb{P}\text{-a.s.} \),

\[
\lim_{N \to \infty} \sqrt{N} \mathbb{E}^{1/p} \left( \left| \phi_X^N \langle Y_0^{n-1} \rangle h - \phi_X \langle Y_0^{n-1} \rangle h \right|^p \bigg| Y_0^{n-1} \right) = \sqrt{2 \sigma^2 \langle Y_0^{n-1} \rangle (h) \left( \frac{\Gamma((p + 1)/2)}{\sqrt{2\pi}} \right)}^{1/p}.
\]
Example: Linear Gaussian HMM

Estimated variance of the bootstrap filter (*):

![Graph showing estimated variance over time steps]

J. Olsson

Long-term stability of SMC methods
1. Hidden Markov models

2. Particle filters
   - The bootstrap particle filter
   - Uniform convergence of particle filters

3. Some novel results
   - Aims
   - Assumptions
   - Main results

4. Elements of proof
Elements of proof

The proof of our main result goes along the following steps.

1. Embed the observation sequence \((Y_n)_{n \geq 0}\) in a two-sided strictly stationary sequence \((Y_n)_{n \in \mathbb{Z}}\).

2. Let \(\pi_{\chi}\langle Y_0^{k-1}\rangle(y_k)\) denote the density of the conditional distribution of \(Y_k\) given \(Y_0^{k-1}\). With this notation,

\[
L_{\chi}\langle Y_0^{n-1}\rangle = \text{Likelihood} = \prod_{k=0}^{n-1} \pi_{\chi}\langle Y_0^{k-1}\rangle(Y_k).
\]
(Douc and Moulines, 2011) For all \( \chi \) and \( \chi' \) there are constants \( C_{\chi,\chi'} > 0 \) and \( \beta \in (0, 1) \) such that, a.s.,

\[
\left| \ln \pi_\chi \langle Y_{-m}^{k-1} \rangle (Y_k) - \ln \pi_{\chi'} \langle Y_{-m}^{k-1} \rangle (Y_k) \right| \leq C_{\chi,\chi'} \beta^{k+m}.
\]

This implies that there exists a function \( \pi \) on the infinite past observation history such that, a.s.,

\[
\lim_{m \to \infty} \pi \langle Y_{-m}^{-1} \rangle (Y_0) = \pi \langle Y_{-\infty}^{-1} \rangle (Y_0).
\]

In addition, a.s.,

\[
\lim_{n \to \infty} \frac{1}{n} \ln L\chi \langle Y_0^{n-1} \rangle = \lim_{n \to \infty} \frac{1}{n} \sum_{k=0}^{n-1} \ln \pi \langle Y_0^{k-1} \rangle (Y_k)
\]

\[
= \lim_{n \to \infty} \frac{1}{n} \sum_{k=0}^{n-1} \ln \pi \langle Y_{-\infty}^{-1} \rangle (Y_k) = \ell_\infty \overset{\Delta}{=} \mathbb{E} \left( \ln \pi \langle Y_{-\infty}^{-1} \rangle (Y_0) \right).
\]

relative entropy
Bound $\sigma^2 \chi \langle Y_0^{n-1} \rangle(h)$ by the product of two quantities, namely

$$
\left( \sup_{(k,m) \in \mathbb{N}^2: k \leq m} \prod_{\ell = k}^{m-1} \frac{\pi \langle Y_{-\infty}^{k-1} \rangle(Y_k)}{\pi \chi \langle Y_0^{k-1} \rangle(Y_k)} \right)^4
$$

which is finite by the exponential forgetting,
Bound $\sigma^2 \langle Y_0^{n-1} \rangle(h)$ by the product of two quantities, namely

$$
\left( \sup_{(k,m) \in \mathbb{N}^2 : k \leq m} \prod_{\ell=k}^{m-1} \frac{\pi \langle Y_{-\infty}^{k-1} \rangle(Y_k)}{\pi \chi \langle Y_0^{k-1} \rangle(Y_k)} \right)^4,
$$

which is finite by the exponential forgetting, and a sum where the $m^{\text{th}}$ term has the same distribution as

$$
\left( L_{\chi \langle Y_{-m}^{-1} \rangle} \times L_{\phi \chi \langle Y_{-m-1}^{-1} \rangle} \langle Y_{-m}^{-1} \rangle \right)
\left( \sup_{x \in X} L_{\delta_x \langle Y_{-m}^{-1} \rangle} \right)^2
\left| \phi \chi \langle Y_{-m}^{-1} \rangle h - \phi \phi \chi \langle Y_{-m-1}^{-1} \rangle \langle Y_{-m}^{-1} \rangle h \right|
$$

uniformly bounded by $C \beta^m$

$$
\times \left( \sup_{x \in X} L_{\delta_x \langle Y_{-m}^{-1} \rangle} \right)^2
\left( \prod_{\ell=1}^{m} \frac{\pi \langle Y_{-\ell}^{\ell-1} \rangle(Y_{-\ell})}{\prod_{\ell=1}^{m} \pi \langle Y_{-\ell}^{\ell-1} \rangle(Y_{-\ell})} \right).
$$
To complete the proof it is enough to show that the quantity

\[
\left( \frac{\sup_{x \in \mathcal{X}} L_{\delta x} \langle Y_{-m}^{-1} \rangle}{\prod_{\ell=1}^{m} \pi \langle Y_{-\infty}^{-\ell-1} \rangle(Y_{-\ell})} \right)^2
= \exp \left( 2m \left[ \frac{1}{m} \ln \sup_{x \in \mathcal{X}} L_{\delta x} \langle Y_{-m}^{-1} \rangle - \frac{1}{m} \sum_{\ell=1}^{m} \ln \pi \langle Y_{-\infty}^{-\ell-1} \rangle(Y_{-\ell}) \right] \right)
\]

grows at most subgeometrically fast with \( m \). This follows however by

\[
\lim_{m \to \infty} \frac{1}{m} \sum_{\ell=1}^{m} \ln \pi \langle Y_{-\infty}^{-\ell-1} \rangle(Y_{-\ell}) = \lim_{m \to \infty} \frac{1}{m} \ln \sup_{x \in \mathcal{X}} L_{\delta x} \langle Y_{-m}^{-1} \rangle = \ell_{\infty},
\]

a result of independent interest.
Summary

We have

- overviewed time uniform convergence of particle filters.
- showed that the asymptotic variance and $L^p$ error of the particle filter is tight under weak assumptions.

Our contribution is twofold, since we

1. present time uniform bounds that also provide the rate of convergence in $\mathcal{N}$ of the particle filter for very general HMMs (with possibly non-compact state space).
2. establish long-term stability of the particle filter also in the case of model misspecification.