Online convergence assessment of particle filters

(Particle filters with adaptive number of particles)

Joaquin Miguez

Department of Signal Theory & Communications,
Universidad Carlos III de Madrid.
E-mail: `joaquin.miguez@uc3m.es`

Joint work with **Victor Elvira** (IMT Lille Douai) and **Petar M. Djurić** (Stony Brook University)

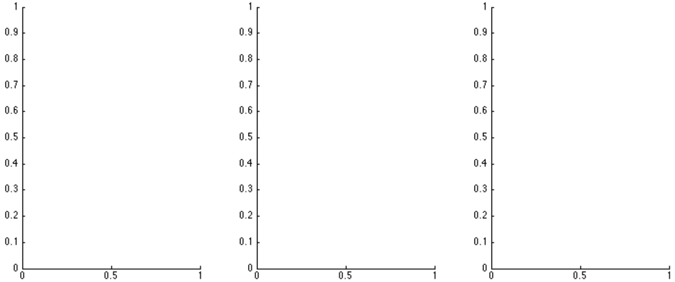
Workshop on Particle Methods & Data Assimilation
Imperial College London – May 9, 2018

Index

Index

Introduction

- Particle filters & their computational cost



- How many particles do we need?
- Can we select this number non-heuristically?
- Can we do it automatically and online?

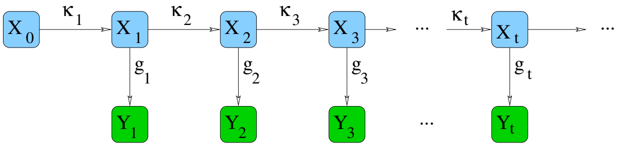
Problem statement

- The performance of a particle filter depends on the ***computational effort***.
  - The error bounds and the practical performance of particle filters depend on the chosen ***number of particles***, N.
  - Small values of N lead to large errors but...
  - ...taking N too large causes delays or the inability to process the observations in real time.
- To determine whether N particles are sufficient or not for a practical application ***we need to be able to assess convergence as we run the PF, typically without knowing the ground truth.***

State of the art

- Early (***heuristic***) work of D. Fox (2002 and 2003): approximating the KLD wrt to a discretised version of the optimal filter
- Straka and Šimandl (2006), effective sample size-based heuristics. No theoretical guarantees.
- Offline selection based on an *ad hoc* AR model (Bhadra and Ionides, 2016)
- ***Mathematically grounded*** work as well...
- The "refueling" approach of J. Cornebise (PhD thesis, 2009)
- Variance estimators (Lee and Whiteley, 2016) for offline allocation of particles.
- Connection with "alive particle filters": random number of particle generation until N "suitable" ones (non-zero weight) are collected (LeGland and Oudjane, 2005; Jasra et al, 2013; Del Moral et al, 2015).
- Similar approach (with a positive lower threshold for likelihood) in Hu et al, 2008.

State space Markov model



- The model

$$
\begin{aligned}
X_0 &\sim \kappa_0(dx) && \text{(prior)} \\
X_t &\sim \kappa_t(dx|x_{t-1}) && \text{(Markov kernel)} \\
Y_t &\sim g_t(y_t|x_t)dy_t && \text{(density of observations)}
\end{aligned}
$$

- $X_t \in \mathcal{X} \subseteq \mathbb{R}^{d_x}$, $t \geq 0$, is the ***unobserved state*** of the system.
- $Y_t \in \mathcal{Y} \subseteq \mathbb{R}^{d_y}$, $t \geq 1$, is the ***sequence of observations***.

Optimal filtering

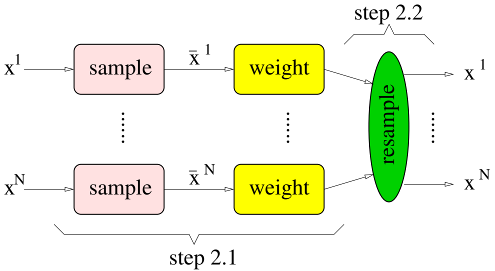
- The model: $X_0 \sim \kappa_0, \quad X_t \sim \kappa_t(\cdot|x_{t-1}), \quad Y_t \sim g_t(\cdot|x_t)$
- We aim at computing the probability measures
  - $\pi_t(A) \triangleq \mathbb{P}\{X_t \in A | Y_{1:t} = y_{1:t}\}$
    (optimal filter)
  - $\xi_t(A') \triangleq \mathbb{P}\{X_t \in A' | Y_{1:t-1} = y_{1:t-1}\}$
    (one-step-ahead predictive measure)
- The filter and the predictive measures satisfy the relationships

$$
\begin{aligned}
\xi_t(dx_t) &= \int \kappa_t(dx_t|x_{t-1})\pi_{t-1}(dx_{t-1}) \quad (\textbf{prediction}) \\
\pi_t(dx_t) &\propto g_t(y_t|x_t)\xi_t(dx_t) \quad (\textbf{update})
\end{aligned}
$$

The standard (bootstrap) particle filter

Algorithm

1. **Initialisation.** *Draw* N *i.i.d. samples* $x_0^{(i)}$, $i = 1, ..., N$, *from* κ_0.
2. **Recursive step.** $\{x_{t-1}^{(i)}\}_{1 \le i \le N}$ *are available at time* $t - 1$.
   2.1 *Draw* $\bar{x}_t^{(i)} \sim \kappa_t(dx_t | x_{t-1}^{(i)})$ *and compute normalised weights* $w_t^{(i)} \propto g_t^{y_t}(\bar{x}_t^{(i)})$ *for* $i = 1, ..., N$.
   2.2 *Draw* $x_t^{(i)} = \bar{x}_t^{(j)}$, $i = 1, ..., N$, *with prob.* $w_t^{(j)}$, $j \in \{1, ..., N\}$.



Approximation of measures & integrals

- The **target** is to approximate **integrals for the form**

$$
\begin{aligned}
(f, \xi_t) &= \int_{\mathcal{X}} f(x)\xi_t(dx) = \mathbb{E}\left[f(X_t)|y_{1:t-1}\right] \\
(f, \pi_t) &= \int_{\mathcal{X}} f(x)\pi_t(dx) = \mathbb{E}\left[f(X_t)|y_{1:t}\right],
\end{aligned}
$$

where $f : \mathcal{X} \to \mathbb{R}$ is bounded.

- **Particle approximations** are straightforward: for the measures we have

$$
\xi_t^N(A') = \frac{1}{N}\sum_{i=1}^{N} \delta_{\bar{x}_t^{(i)}}(A') \quad \text{and} \quad \pi_t^N(A) = \frac{1}{N}\sum_{i=1}^{N} \delta_{x_t^{(i)}}(A),
$$

hence the integrals become

$$
\begin{aligned}
(f, \xi_t) &\approx (f, \xi_t^N) = \int_{\mathcal{X}} f(x)\xi_t^N(dx) = \frac{1}{N}\sum_{i=1}^{N} f(\bar{x}_t^{(i)}), \\
(f, \pi_t) &\approx (f, \pi_t^N) = \int_{\mathcal{X}} f(x)\pi_t^N(dx) = \frac{1}{N}\sum_{i=1}^{N} f(x_t^{(i)}).
\end{aligned}
$$

Index

Prediction measure

- Most of this work revolves around the ability to predict upcoming observations.
- Denote $g_t^{y_t}(x_t) := g_t(y_t|x_t)$ for the pdf of $Y_t|X_t = x_t$.
- The (predictive) pdf of $Y_t|Y_{1:t-1} = y_{1:t-1}$ is

$$p_t(y_t) := (g_t^{y_t}, \xi_t) = \int_{\mathcal{X}} g_t^{y_t}(x_t)\xi_t(dx_t)$$

- The (predictive) probability measure associated to $Y_t|Y_{1:t-1} = y_{1:t-1}$ is

$$\mu_t(dy_t) = p_t(y_t)dy_t$$

- The ***prediction measure*** μ_t ***is absolutely continuous*** wrt the reference measure of $g_t^{y_t}$.

Why is the prediction measure useful?

- We can use μ_t to check out whether "things are working well".
- In the absence of ground truth for the state X_t, we can only test the model by looking at the observations.
  - If the filter is (nearly) exact, ***we can validate*** whether ***the model*** is capable of predicting the observations (Djurić and M., 2010).
  - If the model is assumed valid, ***we can test*** whether ***the filtering algorithm*** is yielding accurate results (Elvira et al, 2017).
  - If the model and the filter are both assumed valid, ***we can test for outliers*** (Maíz et al, 2012).
- We focus on the second case: we assume a valid model and ***tune the effort of the particle filter*** until a desired level of accuracy is attained.

Particle approximation of the prediction measure

- The particle filter enables the approximation of $p_t(y_t)$ and $\mu_t(dy_t)$.
- Approximation of the prediction pdf

$$p_t^N(y_t) := (g_t^{y_t}, \xi_t^N) = \frac{1}{N}\sum_{i=1}^{N} g_t^{y_t}(\bar{x}_t^{(i)}).$$

- Standard theory yields

$$\lim_{N\to\infty} |p_t(y) - p_t^N(y)| = 0 \quad \text{a.s., with errors } \mathcal{O}(N^{-\frac{1}{2}+\epsilon}),$$

  but only point-wise (e.g., Del Moral, 2004), not uniformly over $\mathcal{Y}$.
- The prediction measure can be approximated as

$$\mu_t^N(dy) := p_t^N(y)dy$$

- This is a random measure, continuous wrt the reference measure of $g(y|x)$. ***Not a classical point-mass Monte Carlo approximation.*** How does it converge?

Assumptions on the model

($\mathfrak{L}$) The likelihood $g_t^{y_t}$ is positive and bounded,

$$\|g_t\|_\infty := \sup_{x\in\mathcal{X}, y\in\mathcal{Y}} g_t^y(x) < \infty.$$

($\mathfrak{D}$) The likelihood $g_t^y(x)$ is differentiable wrt y, with bounded derivatives up to order d_y, i.e.,

$$\|D^{\mathbf{1}} g_t\|_\infty := \sup_{x\in\mathcal{X}, y\in\mathcal{Y}} |D^{\mathbf{1}} g_t^y(x)| < \infty,$$

where $D^{\mathbf{1}} g_t^y(x) = \frac{\partial^{d_y} g_t}{\partial y_{\mathbf{1}} \dots \partial y_{d_y}}$.

($\mathfrak{C}$) For any $0 < \beta < 1$ and any $p \geq 4$, the sequence

$$C_N := \left[-\frac{N^{\frac{\beta}{p}}}{2}, +\frac{N^{\frac{\beta}{p}}}{2} \right]^{d_y} \quad \text{satisfies the inequality } \mu_t(\overline{C_N}) \leq bN^{-\eta}$$

for some constants $b > 0$ and $\eta > 0$ independent of N (with $\overline{C_N}$ denoting the complement $\mathcal{X} \backslash C_N$).

A convergence result for μ_t^N

Recall the constructions $p_t^N(y) \triangleq (g_t(y|\cdot), \xi_t^N)$ and $\mu_t^N(dy) = p_t^N(y)dy$.

Theorem (1)

If assumptions $(\mathfrak{L})$, $(\mathfrak{D})$ *and* $(\mathfrak{C})$ *hold, then for every bounded function* $f : \mathcal{X} \to \mathbb{R}$ *and any* $\epsilon \in (0, \frac{1}{2})$ *there exists an a.s. finite random variable* W_t^ϵ*, independent of* N*, such that*

$$\left|(f, \mu_t^N) - (f, \mu_t)\right| \leq \frac{W_t^\epsilon}{N^{(\frac{1}{2}-\epsilon)\wedge\eta}}.$$

In particular, $\lim_{N\to\infty}(f, \mu_t^N) = (f, \mu_t)$ *a.s.*

A convergence result for μ_t^N

- The argument of the proof is based on the analysis of kernel density estimators of filtering pdf's in (Crisan and M., 2014).
- A key result:
  - ***Uniform convergence*** of $p_t^N(y)$: under assumptions $(\mathfrak{L})$, $(\mathfrak{C})$ and $(\mathfrak{D})$ the approximation p_t^y converges uniformly on the compacts C_N,

$$\sup_{y \in C_N} |p_t(y) - p_t^N(y)| \leq \frac{V_t^\epsilon}{N^{\frac{1}{2}-\epsilon}}$$

    for any $\epsilon \in (0, \frac{1}{2})$ and some a.s. finite r.v. V_t^ϵ independent of N.
  - Hence

$$\lim_{N\to\infty} C_N = \mathcal{X} \quad \text{and} \quad \lim_{N\to\infty} \sup_{y \in C_N} |p_t(y) - p_t^N(y)| = 0.$$

A remark on μ_t^N

- The particle approximation of the prediction measure, μ_t^N, is not useful to numerically compute integrals $(f, \mu_t) \approx (f, \mu_t^N)$.
- It is just a ***theoretical tool*** that enables the ***analysis of*** expressions involving synthetically-generated, ***fictitious observations*** $\tilde{Y}_t \sim \mu_t^N$ and compare them with real observations Y_t (under the hypothesis $Y_t \sim \mu_t$).

Index

Invariant predictive statistics

- The gist: assess performance of the particle filter online by ***testing how well it can predict the next observation.*** Assume scalar observations ($d_y = 1$) in the sequel.
- ***Assume*** that $Y_t | Y_{1:t-1} = y_{1:t-1} \sim \mu_t$ (i.e., our model represents well the mechanism that generates the observations) and let y_t be the actual observation at time t.
- Draw J iid fictitious observations $\bar{y}_t^{(1)}, \ldots, \bar{y}_t^{(J)}$ from μ_t and build the set

$$\mathcal{A}_{t,J} \triangleq \{y \in \{\bar{y}_t^{(j)}\}_{1 \leq j \leq J} : y < y_t\}.$$

- The r.v. $A_{t,J} \triangleq |\mathcal{A}_{t,J}|$ represents the number of fictitious observations smaller than the actual one, y_t.
- The distribution of the r.v. $A_{t,J}$ is uniform no matter the state space model.

Invariant predictive statistics

- Recall the set

$$\mathcal{A}_{t,J} \triangleq \{y \in \{\bar{y}_t^{(j)}\}_{1 \leq j \leq J} : y < y_t\}$$

and let $A_{t,J} = |\mathcal{A}_{t,J}|$, the number fictitious observations $\bar{y}_t^{(j)} < y_t$.

Proposition

If $y_t, \bar{y}_t^{(1)}, \ldots, \bar{y}_t^{(J)} \sim \mu_t$ iid, then the probability mass function (pmf) of the r.v. $A_{t,J}$, denoted $\mathbb{Q}_J$, is uniform over the set $\{0, 1, \ldots, J\}$, namely

$$\mathbb{Q}_J(n) = \frac{1}{J+1}, \quad 0 \leq n \leq J.$$

- We can exploit this invariance property to check convergence!

(Approximately) Invariant predictive statistics

- The PF yields estimates of the predictive distribution of Y_t, namely

$$p_t^N(y) \triangleq (g_t(y|\cdot), \xi_t^N) = \frac{1}{N}\sum_{i=1}^{N} g_t(y|\bar{x}_t^{(i)}) \quad \text{and} \quad \mu_t^N(dy) = p_t^M(y)dy$$

- Let $\{\tilde{y}_t^{(k)}\}_{k=1}^J$ be iid draws from μ_t^N. We introduce the set $\mathcal{A}_{t,J,N} := \{y \in \{\tilde{y}_t^{(j)}\}_{j=1}^J : y < y_t\}$ and the associated r.v. $A_{t,J,N} = |\mathcal{A}_{t,J,N}|$ with pmf $\mathbb{Q}_{t,J,N}$. Then...

Theorem (2)

If the history of observations $y_{1:t-1}$ is fixed and assumptions $(\mathfrak{L})$, $(\mathfrak{D})$ and $(\mathfrak{D})$ hold, then there exist non-negative r.v.'s $\{\varepsilon_t^N\}_{N\in\mathbb{N}}$ such that $\lim_{N\to\infty} \varepsilon_t^N = 0$ a.s. and

$$\frac{1}{J+1} - \varepsilon_t^N \leq \mathbb{Q}_{t,J,N}(n) \leq \frac{1}{J+1} + \varepsilon_t^N.$$

In particular, $\lim_{N\to\infty} \mathbb{Q}_{t,J,N}(n) = \mathbb{Q}_J(n) = \frac{1}{J+1}$ a.s.

We have got an asymptotically invariant statistic that we can compute easily from the particle filter.

A quick summary up to here...

1. We can compute a particle approximation of the prediction measure, $\mu_t^N \stackrel{N\to\infty}{\longrightarrow} \mu_t$.
2. From μ_t^N, we can draw (a few) fictitious observations $\tilde{y}_t^{(1)}, \ldots, \tilde{y}_t^{(J)}$.
3. When we collect y_t, the actual observation, we let $A_{t,J,N}$ be the number of fictitious observations smaller than y_t.
4. $A_{t,J,N}$ is asymptotically uniform (as $N \to \infty$).

A quick summary up to here...

1. We can compute a particle approximation of the prediction measure, $\mu_t^N \stackrel{N\to\infty}{\longrightarrow} \mu_t$.
2. From μ_t^N, we can draw (a few) fictitious observations $\tilde{y}_t^{(1)}, \ldots, \tilde{y}_t^{(J)}$.
3. When we collect y_t, the actual observation, we let $A_{t,J,N}$ be the number of fictitious observations smaller than y_t.
4. $A_{t,J,N}$ is asymptotically uniform (as $N \to \infty$).

A quick summary up to here...

1. We can compute a particle approximation of the prediction measure, $\mu_t^N \overset{N\to\infty}{\longrightarrow} \mu_t$.
2. From μ_t^N, we can draw (a few) fictitious observations $\tilde{y}_t^{(1)}, \ldots, \tilde{y}_t^{(j)}$.
3. When we collect y_t, the actual observation, we let $A_{t,J,N}$ be the number of fictitious observations smaller than y_t.
4. $A_{t,J,N}$ is asymptotically uniform (as $N \to \infty$).

A quick summary up to here...

1. We can compute a particle approximation of the prediction measure, $\mu_t^N \stackrel{N\to\infty}{\longrightarrow} \mu_t$.
2. From μ_t^N, we can draw (a few) fictitious observations $\tilde{y}_t^{(1)}, \ldots, \tilde{y}_t^{(j)}$.
3. When we collect y_t, the actual observation, we let $A_{t,J,N}$ be the number of fictitious observations smaller than y_t.
4. $A_{t,J,N}$ is asymptotically uniform (as $N \to \infty$).

An algorithm with adaptive number of particles

Choose...

- Bounds for the number of particles, $\underline{N} \leq N_k \leq \overline{N}$.
- Functions $f_-(N) < N$, to decrease the number of particles, ...
- ... and $f_+(N) > N$, to increase the number of particles.
- A window size $W > 1$, to collect and assess the statistic $A_{t,J,N}$.
- Two thresholds $0 < p_\ell < p_h < 1$.

An algorithm with adaptive number of particles

1. **[Initialization]**

   1.1 Draw $x_0^{(i)} \sim \kappa_0$ for $i = 1, ..., N_0$. Set $w_0^{(i)} = \frac{1}{N_0}$ $\forall i$.

   and set $n = 1$.

2. **[For $t = 1 : T$]**

   2.1 **Standard particle filtering:** prediction, update & resampling with N_k particles.

   2.2 **Fictitious observations:**
   - Draw $\tilde{y}_t^{(j)} \sim \mu_t^{N_k}, j = 1, \ldots, J$.
   - Compute A_{t,J,N_k} (no. of artificial observations smaller than y_t).

   2.3 If $t = W \times k$, (**assessment of convergence**):
   - Compute the χ_t^2 statistic over the empirical distribution of $A_{t-W+1:t,J,N_k}$.
   - Calculate the p-value $p_{J,t}^*$ by comparing the statistic χ_t^2 to the χ^2-distribution with J degrees of freedom.
   - If $p_{J,t}^* \leq p_\ell$, increase $N_k = f_+(N_{k-1}) \wedge \overline{N}$.
   - Else, if $p_{J,t}^* \geq p_h$, decrease $N_k = f_-(N_{k-1}) \vee \underline{N}$.
   - Else, $N_k = N_{k-1}$.
   - Set $k = k + 1$.

Testing for convergence

1. **[Initialization]**

   1.1 Draw $x_0^{(i)} \sim \kappa_0$ for $i = 1, ..., N_0$. Set $w_0^{(i)} = \frac{1}{N_0}$ $\forall i$.

   and set $n = 1$.

2. **[For $t = 1 : T$]**

   2.1 **Standard particle filtering:** prediction, update & resampling with N_k particles.

   2.2 **Fictitious observations:**
   - Draw $\tilde{y}_t^{(j)} \sim \mu_t^{N_k}, j = 1, \ldots, J$.
   - Compute A_{t,J,N_k} (no. of artificial observations smaller than y_t).

   2.3 If $t = W \times k$, (**assessment of convergence**):
   - Compute the χ_t^2 statistic over the empirical distribution of $A_{t-W+1:t,J,N_k}$.
   - Calculate the p-value $p_{J,t}^*$ by comparing the statistic χ_t^2 to the χ^2-distribution with J degrees of freedom.
   - If $p_{J,t}^* \leq p_\ell$, increase $N_k = f_+(N_{k-1}) \wedge \overline{N}$.
   - Else, if $p_{J,t}^* \geq p_h$, decrease $N_k = f_-(N_{k-1}) \vee \underline{N}$.
   - Else, $N_k = N_{k-1}$.
   - Set $k = k + 1$.

Updating the number of particles

1. **[Initialisation]**

   1.1 Draw $x_0^{(i)} \sim \kappa_0$ for $i = 1, ..., N_0$. Set $w_0^{(i)} = \frac{1}{N_0}$ $\forall i$.

   and set $n = 1$.

2. **[For $t = 1 : T$]**

   2.1 **Standard particle filtering:** prediction, update & resampling with N_k particles.

   2.2 **Fictitious observations:**
   - Draw $\tilde{y}_t^{(j)} \sim \mu_t^{N_k}, j = 1, \ldots, J$.
   - Compute A_{t,J,N_k} (no. of artificial observations smaller than y_t).

   2.3 If $t = W \times k$, (**assessment of convergence**):
   - Compute the χ_t^2 statistic over the empirical distribution of $A_{t-W+1:t,J,N_k}$.
   - Calculate the p-value $p_{J,t}^*$ by comparing the statistic χ_t^2 to the χ^2-distribution with J degrees of freedom.
   - If $p_{J,t}^* \leq p_\ell$, increase $N_k = f_+(N_{k-1}) \wedge \overline{N}$.
   - Else, if $p_{J,t}^* \geq p_h$, decrease $N_k = f_-(N_{k-1}) \vee \underline{N}$.
   - Else, $N_k = N_{k-1}$.
   - Set $k = k + 1$.

Stochastic Lorenz 63 model

- Continuous-time SDE:

$$\begin{aligned} dX_1 &= -s(X_1 - Y_1) + dW_1, \\ dX_2 &= rX_1 - X_2 - X_1X_3 + dW_2, \\ dX_3 &= X_1X_2 - bX_3 + dW_3, \end{aligned}$$

- Parameter values $(s, r, b) = \left(10, 28, \frac{8}{3}\right)$.
- Discretised via the Euler-Maruyama scheme with time-step $\Delta = 10^{-3}$.
- Observations:

$$Y_n = X_1(200\Delta n) + V_n, \quad n = 1, 2, \ldots$$

with $V_n \sim \mathcal{N}(0, \sigma^2)$, $\sigma^2 = 0.5$.

Performance for different thresholds

$[p_l - p_h]$	Fixed $N = 2^{15}$	$[0.4 - 0.8]$	$[0.35 - 0.7]$	$[0.3 - 0.7]$	$[0.25 - 0.65]$	$[0.2 - 0.6]$
MSE	1.5193	1.5234	1.5240	1.5287	3.7552	4.6540
$\hat{N}$	32768	24951	14840	8729	2197	451
p-val	0.5108	0.5089	0.4902	0.4815	0.4872	0.4785
Hell. distance	0.2312	0.2355	0.2493	0.2462	0.2476	0.2521
exec. time (s)	6201	5617	3014	1532	131	67
time ratio	1	1.10	2.1	4.05	47.43	92.36

Table: Lorenz 63 Model: $\Delta = 10^{-3}$, $T_{obs} = 200\Delta$, $\sigma^2 = 0.5$. Algorithm details: window size $W = 20$, $J = 7$ artificial observations, range of N_k given by $\overline{N} = 2^{15}$, $\underline{N} = 2^7$. MSE in the approximation of the posterior mean, average number of particles $\hat{N}$, averaged p-value, and averaged Hellinger distance.

Stabilisation of N_k

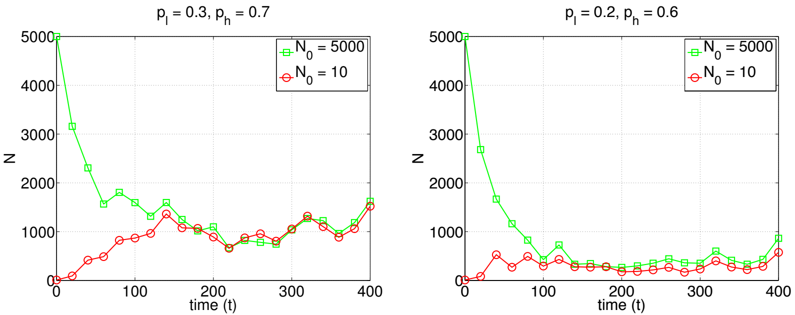


Figure: Lorenz 63 Model. Evolution of the number of particles adapted by the proposed algorithm when the initial number of particles $N_0 \in \{10, 5000\}$.

Index

Matching the uniform statistics

- So far we have seen sufficient conditions to justify the us of the invariant statistic $A_{t,J,N}$: if the filter is doing its job, then $A_{t,J,N}$ is approximately uniform.
- The natural question is what can be said in the opposite direction: if $A_{t,J,N}$ is uniform, what can we claim about the particle approximations?
- Recall that if f is a pdf and $\mathsf{F}(y) = \int_{-\infty}^{y} \mathsf{f}(z)dz$ is the associated cumulative distribution function (cdf) then

$$(\mathsf{F}^m, \mathsf{f}) = \int_{-\infty}^{\infty} \mathsf{F}(y)^m \mathsf{f}(y)dy = \frac{1}{m+1}.$$

Matching the uniform statistics

- Let $F_t(y) = \int_{-\infty}^{\infty} p_t(y)dy$ the exact prediction cdf.
- Let $F_t^N(y) = \int_{-\infty}^{\infty} p_t^N(y)dy$ the approximate prediction cdf.

Proposition

If the r.v. $A_{t,J,N}$ is uniform then

$$((F_t^N)^m, p_t) = (F_t^m, p_t) = \frac{1}{m+1}, \quad \textit{for } m = 0, 1, \ldots, J.$$

- This means that the approximate cdf $F_t^N(y) = \mu_t^N((-\infty, y])$ is "as good as" the actual $F_t(y)$ for moments up to order J, where J is the number of fictitious observations.

A block-adaptive particle filter

- One more sanity check: we have
  - shown that we can assess the convergence of a particle filter online,
  - and used this property to update the number of particles.
- When $N_k < N_{k-1}$ the error bound increases ($N_k^{-\frac{1}{2}} > N_{k-1}^{-\frac{1}{2}}$), so we can expect the performance to deteriorate.
- However, can we ensure that the performance actually improves when $N_k > N_{k-1}$?? This is not directly implied by standard theory.
- The answer is ***yes, for uniformly convergent particle filters.***

A general block-adaptive bootstrap filter

Variable window size W_k and number of particles N_k.

1. **[Initialisation]**
   1.1 Draw $x_0^{(m)}$ iid from the prior κ_0, assign $w_0^{(m)} = 1/M_0$, $m = 1, \ldots, M_0$.
   1.2 Set $k = 0$ (block counter) and $W_0 > 0$ (size of the first window).
2. **[For $t = 1, 2, ...$]**
   2.1 **Bootstrap particle filter:**
   - Propagate the particles $\bar{x}_t^{(i)} \sim \kappa_t(dx_t|x_{t-1}^{(i)})$, $i = 1, \ldots, N_k$.
   - Compute normalised weights $\bar{w}_t^{(i)} \propto g_t(\bar{x}_t^{(i)})$, $i = 1, \ldots, N_k$.

   2.2 **Fictitious observations:**
   - Draw $\tilde{y}_t^{(j)} \sim p_t^{N_k}(y_t), j = 1, \ldots, J$.
   - Compute $A_{t,J,N_k} = a_{t,J,N_k}$.

   2.3 **Assessment of convergence:** If $t = \sum_{r=0}^{k} W_r - 1$ then:
   - Test the sequence $a_{t-W_k+1:t,J,N_k}$.
   - Set $k = k + 1$. Update the number of particles $N_k > 0$ and the window size $W_k > 0$..
   - Resample N_k particles with replacement.

   Else:
   - Resample N_k particles with replacement.

Assumptions for the analysis

- Let $\mathcal{P}(\mathcal{X})$ the set of probability measures on $(\mathcal{B}(\mathcal{X}), \mathcal{X})$. We introduce the operators $\psi_t : \mathcal{P}(\mathcal{X}) \to \mathcal{P}(\mathcal{X})$ satisfying the relationships

$$(f, \psi_t(\alpha)) = \frac{(g_t^{y_t}(f, \kappa_t), \alpha)}{(g_t, \alpha)}, \quad t = 1, 2, ...,$$

  for any $\alpha \in \mathcal{P}(\mathcal{X})$. Then $\pi_t = \psi_t(\pi_{t-1})$.
- Composition of operators: $\psi_{t|t-r}(\alpha) := (\psi_t \circ \psi_{t-1} \circ \cdots \circ \psi_{t-r+1})(\alpha)$. It is apparent that $\pi_t = \psi_{t|t-r}(\pi_{t-r})$.
- Assumptions:
  - (S) For every $f \in B(\mathcal{X})$ and every $r > 0$ there exists $\varepsilon(f, r)$ such that

$$\sup_{\alpha,\beta \in \mathcal{P}(\mathcal{X})} \left| (f, \psi_{t|t-r}(\alpha)) - (f, \psi_{t|t-r}(\beta)) \right| \leq \varepsilon(f, r) \stackrel{r\to\infty}{\longrightarrow} 0.$$

  - (G) There exists a positive constant $a < \infty$ such that

$$0 < \frac{1}{a} < g_t^{y_t}(x) < a < \infty$$

    for every $t \geq 1$ and every $x \in \mathcal{X}$.

Block-adaptive error bounds

Theorem (3)

Let $t_k = \sum_{j=0}^{k} W_j - 2$ and let $\pi_{t_k}^{N_k}$ be the particle approximation of the filtering measure π_{t_n} produced by the block-adaptive bootstrap filter. If assumptions (S) and (G) hold, then for any $0 < \epsilon < \frac{1}{2}$ there exists $W_k = \mathcal{O}(\epsilon \log(N_k))$ such that

$$\left\| (f, \pi_{t_k}^{N_k}) - (f, \pi_{t_k}) \right\|_p < \frac{c\|f\|_\infty}{N_k^{\frac{1}{2}-\epsilon}} + \tilde{\varepsilon}(f, N_k) \quad \textit{and} \quad \lim_{N_k \to \infty} \tilde{\varepsilon}(f, N_k) = 0$$

for every $f \in B(\mathcal{X})$, every $k \geq 1$ and a constant $c < \infty$ independent of ϵ, t_k and N_k.

Remark: the window size can made constant, $W_k = W$. In that case, it can be chosen to make $\tilde{\varepsilon}(f, N_k) = \varepsilon(f, W_k) = \varepsilon(f, W)$ small enough.

Computer simulations

M_1 / M_2	100	1000	100 / 1000	1000	10000	1000 / 10000
MSE (last $T/4$)	$8.90\,10^{-3}$	$9.02\,10^{-4}$	$8.99\,10^{-4}$	$9.02\,10^{-4}$	$8.93\,10^{-5}$	$8.69\,10^{-5}$

Table: Linear Gaussian model: $T = 1000$, $\sigma_x^2 = 0.5$, $\sigma_y^2 = 1$, $a = 0.9$. M_1 particles for $t \in \{1, ..., \frac{T}{2}\}$ and M_2 particles for $t \in \{\frac{T}{2} + 1, ..., T\}$

.

M_1 / M_2	50	1000	50 / 1000	200	4000	200 / 4000	1000	20000	1000 / 20000
MSE (last $T/4$)	16.69	1.493	1.564	4.815	1.386	1.374	1.494	1.348	1.335

Table: Stochastic growth model: $T = 1000$, $\sigma_x = 1$, $\sigma_y = 0.1$, $\phi = 0.4$. M_1 particles for $t \in \{1, ..., \frac{T}{2}\}$ and M_2 particles for $t \in \{\frac{T}{2} + 1, ..., T\}$

.

Index

Conclusions

- Online assessment of the convergence of particle filters using the prediction measure (of the observations).
- Application to devise particle filters with adaptive number of particles.
- Key ingredients:
  - Particle approximation of the prediction measure.
  - Model-invariante statistics.
- Future work:
  - Comparison of statistics & adaptation criteria.
  - Extensions to vector observations.

Index

References

- Djuric, P. M., & Miguez, J. (2010). Assessment of nonlinear dynamic models by Kolmogorov–Smirnov statistics. IEEE Transactions on Signal Processing, 58(10), 5069-5079.
- Maiz, C. S., Molanes-Lopez, E. M., Miguez, J., & Djuric, P. M. (2012). A particle filtering scheme for processing time series corrupted by outliers. IEEE Transactions on Signal Processing, 60(9), 4611-4627.
- Elvira, V., Miguez, J., & Djuric, P. M. (2017). Adapting the number of particles in sequential Monte Carlo methods through an online scheme for convergence assessment. IEEE Transactions on Signal Processing, 65(7), 1781-1794.
- Crisan, D., & Miguez, J. (2014). Particle-kernel estimation of the filter density in state-space models. Bernoulli, 20(4), 1879-1929.

References

- Fox, D. (2002). KLD-sampling: Adaptive particle filters. In Advances in neural information processing systems (pp. 713-720).
- Fox, D. (2003). Adapting the sample size in particle filters through KLD-sampling. The international Journal of robotics research, 22(12), 985-1003.
- LeGland, F., & Oudjane, N. (2005, September). A sequential particle algorithm that keeps the particle system alive. In Signal Processing Conference, 2005 13th European (pp. 1-4). IEEE.
- Cornebise, J. (2009). Adaptive Sequential Monte Carlo Methods (Doctoral dissertation, Ph. D. thesis, University Pierre and Marie Curie–Paris 6).
- Moral, P. D., Jasra, A., Lee, A., Yau, C., & Zhang, X. (2015). The alive particle filter and its use in particle Markov chain Monte Carlo. Stochastic Analysis and Applications, 33(6), 943-974.