

ON THE ASYMPTOTIC VARIANCE IN THE CENTRAL LIMIT THEOREM FOR PARTICLE FILTERS

BENJAMIN FAVETTO¹

Abstract. Particle filter algorithms approximate a sequence of distributions by a sequence of empirical measures generated by a population of simulated particles. In the context of Hidden Markov Models (HMM), they provide approximations of the distribution of optimal filters associated to these models. For a given set of observations, the behaviour of particle filters, as the number of particles tends to infinity, is asymptotically Gaussian, and the asymptotic variance in the central limit theorem depends on the set of observations. In this paper we establish, under general assumptions on the hidden Markov model, the tightness of the sequence of asymptotic variances when considered as functions of the random observations as the number of observations tends to infinity. We discuss our assumptions on examples and provide numerical simulations.

Mathematics Subject Classification. 60G35, 62M20, 60F05, 60J05.

Received November 23, 2009. Received May 4, 2010.

1. INTRODUCTION

Hidden Markov models (or state-space models) form a class of stochastic models which are used in numerous fields of applications. In these models, a discrete time process $(Y_n, n \geq 0)$ – the signal – is observed while the process of interest $(X_n, n \geq 0)$ – the state process – is not observed. The standard assumptions for the joint-process $(X_n, Y_n)_{n \geq 0}$ are that (X_n) is a Markov chain, that, given $(X_n, n \geq 0)$ the random variables (Y_n) are conditionally independent and the conditional distribution of Y_n only depends on the corresponding state variable X_n . For general references, see *e.g.* [13] or [2].

Nonlinear filtering is concerned with the estimation of X_k or the prediction of X_{k+1} given the observations $(Y_0, \dots, Y_k) := Y_{0:k}$. For this, one has to compute the conditional distributions $\pi_{k|k:0} = \mathcal{L}(X_k | Y_k, \dots, Y_0)$ or $\eta_{k+1|k:0} = \mathcal{L}(X_{k+1} | Y_k, \dots, Y_0)$ which are derived recursively by a sequence of measure-valued operators depending on the observations

$$\pi_{k|k:0} = \Psi_{Y_k}(\pi_{k-1|k-1:0}) \text{ and } \eta_{k+1|k:0} = \Phi_{Y_k}(\eta_{k|k-1:0}).$$

For more details, see *e.g.* [6] or [8].

Keywords and phrases. Hidden Markov Model, Particle filter, Central Limit Theorem, Asymptotic variance, Tightness, Sequential Monte-Carlo.

¹ Laboratoire MAP5, Université Paris Descartes, U.F.R. de Mathématique et Informatique, CNRS UMR 8145, 45, rue des Saints-Pères, 75270 Paris Cedex 06, France; Benjamin.Favetto@mi.parisdescartes.fr

Unfortunately, except for very few models, such as the Kalman filter or some other models (for instance, those presented in [3]), these recursions rapidly lead to intractable computations and exact formulae are out of reach. Moreover, the standard Monte-Carlo methods fail to provide good approximations of these distributions (see *e.g.* the introduction in [17]). This justifies the huge popularity of sequential Monte-Carlo methods which are generally the only possible computing approach to solve these problems (see [12] or [16]). Sequential Monte-Carlo methods (or particle filters, or Interacting Particle Systems) are iterative algorithms based on simulated “particles” which provide approximations of the conditional distributions involved in prediction and filtering.

Denoting by $\pi_{k|k:0}^N$ (resp. $\eta_{k+1|k:0}^N$) the particle filter approximations of $\pi_{k|k:0}$ (resp. $\eta_{k+1|k:0}$) based on N particles, several recent contributions have been concerned with the evaluation of errors between the approximate and the exact filter as N grows to infinity, for a given (fixed) set of data (Y_k, \dots, Y_0) (see *e.g.* [10]). In particular, for the bootstrap particle filter, [8] prove that, for a wide class of real-valued functions f , $\sqrt{N}(\pi_{k|k:0}^N(f) - \pi_{k|k:0}(f))$ (resp. $\sqrt{N}(\eta_{k+1|k:0}^N(f) - \eta_{k+1|k:0}(f))$) converges in distribution to $\mathcal{N}(0, \Gamma_{k|k:0}(f))$ (resp. $\mathcal{N}(0, \Delta_{k+1|k:0}(f))$). Central limit theorems for an exhaustive class of sequential Monte-Carlo methods are also proved in [4] and [14].

To our knowledge, still little attention has been paid to the time behaviour (with respect to k) of the approximations. Recently, [17] has studied a uniform time average consistency of Monte-Carlo particle filters (see also [15] for a uniform approximation).

In this paper, we are concerned with the tightness of the asymptotic variances $\Gamma_{k|k:0}(f)$, $\Delta_{k+1|k:0}(f)$ in the central limit theorem for the bootstrap particle filter, when considered as random variables functions of Y_0, \dots, Y_k as $k \rightarrow \infty$. This is an important issue since these asymptotic variances measure the accuracy of the numerical method and provide confidence intervals. In [4], for the case of the bootstrap filter, the asymptotic variance $\Gamma_{k|k:0}(f)$ is proved to be bounded from above by a constant, under stringent assumptions on the conditional distribution of Y_i given X_i and on the transition densities of the unobserved Markov chain. In [9] the asymptotic variance $\Gamma_{k|k:0}(f)$ is proved to be tight (in k) in the case of the Kalman filter. The proof is based on explicit computations which are possible in this model. Below, we consider a general model and prove the tightness of both $\Gamma_{k|k:0}(f)$ and $\Delta_{k+1|k:0}(f)$ for f a bounded function under a set of assumptions which are milder than those in [4] but which do not include the Kalman filter. In general, authors concentrate on filtering rather than on prediction as filtering is more important for applications. However, from the theoretical point of view, we focus on the prediction because computations are a little simpler. First we prove the tightness of the asymptotic variances $\Delta_{k+1|k:0}(f)$ obtained in the central limit theorem for prediction, and then we deduce the analogous result for $\Gamma_{k|k:0}(f)$. For the transition kernel of the Markov chain, we rely on a strong assumption, which mainly holds when the state space of the hidden chain is compact (Assumption (A)). Nevertheless, such an assumption is of common use in this kind of studies (see *e.g.* [1], [10]). In the sense of [10], it means that the whole state space of the hidden chain is “small” (see [10]). On the other hand, our assumptions on the conditional distributions of Y_i given X_i are standard ((B1)–(B2)). Assumption (B3) involves the distribution of the observations, and is easy to check on several classical models.

The paper is organized as follows. In Section 2, we present our notations and assumptions, and give the formulae for $\Gamma_{k|k:0}(f)$ and $\Delta_{k+1|k:0}(f)$ and some preliminary propositions in order to obtain formulae as simple as possible for the asymptotic variances. Section 3 is devoted to the proof of the tightness of $\Delta_{k+1|k:0}(f)$ from which we deduce the tightness of $\Gamma_{k|k:0}(f)$. Moreover, we illustrate our assumptions on examples and provide some numerical simulation results.

2. NOTATIONS, ASSUMPTIONS AND PRELIMINARY RESULTS

Let (X_k) be the time-homogeneous hidden Markov chain, with state space \mathcal{X} and transition kernel $Q(x, dx')$. The observed random variables (Y_k) take values in another space \mathcal{Y} and are conditionally independent given $(X_k)_{k \geq 0}$ with $\mathcal{L}(Y_i | (X_k)_{k \geq 0}) = F(X_i, dy)$. For $0 \leq i \leq k$, denote by $Y_{i:k}$ the vector $(Y_i, Y_{i+1} \dots Y_k)$. Denote also, for f a bounded measurable function, $Qf(x) = \int f(x')Q(x, dx')$.

Denote by $\pi_{k|k:0}^{\eta_0} = \mathcal{L}_{\eta_0}(X_k|Y_{k:0})$ (resp. $\eta_{k|k-1:0}^{\eta_0} = \mathcal{L}_{\eta_0}(X_k|Y_{k-1:0})$) the filtering distribution (resp. the predictive distribution) at step k when η_0 is the initial distribution of the chain (distribution of X_0). By convention, $\eta_{0|-1:0}^{\eta_0} = \eta_0$.

Let us now introduce our assumptions. Assumptions A concern the hidden chain, Assumptions B concern the conditional distribution of Y_i given X_i .

- (A0) \mathcal{X} is a convex subset of \mathbb{R}^d . The transition operator Q admits transition densities with respect to the Lebesgue measure on \mathcal{X} denoted by $dx' : Q(x, dx') = p(x, x')dx'$. The transition densities are positive and continuous on $\mathcal{X} \times \mathcal{X}$. For φ bounded and continuous on \mathcal{X} , $Q\varphi$ is bounded and continuous on \mathcal{X} (Q is Feller).
- (A1) The transition operator Q admits a stationary distribution $\pi(dx)$ having a density h with respect to dx which is continuous and positive on \mathcal{X} .
- (A2) There exists a probability measure μ and two positive numbers $\epsilon_- \leq \epsilon_+$ such that

$$\forall x \in \mathcal{X}, \forall B \in \mathcal{B}(\mathcal{X}) \quad \epsilon_- \mu(B) \leq Q(x, B) \leq \epsilon_+ \mu(B).$$

Moreover, for all f continuous and positive on \mathcal{X} , $\mu(f) > 0$.

- (B1) The conditional distribution of Y_k given X_k has density $f(y|x)$ with respect to a dominating measure $\kappa(dy)$, and $(x, y) \mapsto f(y|x)$ is measurable and positive.
- (B2) $x \mapsto f(y|x)$ is continuous and bounded from above for all y κ a.e.

Under (B2), $q(y) = \sup_{x \in \mathcal{X}} f(y|x)$ is well defined and positive. Up to changing $\kappa(dy)$ into $\frac{1}{q(y)}\kappa(dy)$, we can assume without loss of generality that

$$\forall x \in \mathcal{X}, \quad f(y|x) \leq 1. \tag{2.1}$$

Finally, we introduce an assumption involving the distribution of the observations. Define $g_k(x) := f(Y_k|x)$ for $k \geq 0$.

- (B3) For some $\delta > 0$

$$\sup_{k \geq 0} \mathbf{E} \left| \log \left(\eta_{k|k-1:0}^{\eta_0} (g_k) \right) \right|^{1+\delta} < \infty, \tag{2.2}$$

where \mathbf{E} denotes the expectation with respect to the distribution of $(Y_k)_{k \geq 0}$.

Except (A2) these assumptions are weak and standard. For instance, (A0)–(A1) easily hold for discretized diffusion processes with constant discretization step. Assumption (A2), which is the most stringent, is nevertheless classical and is verified when \mathcal{X} is compact. (see [1] and the chronological discussion in [11]). Assumptions (B1)–(B2) are mild. Note that they are weaker than the corresponding ones in [4] and the same as in [17]. By (A0), for φ non null, non negative and continuous on \mathcal{X} , $Q\varphi > 0$. With (B2), for all y κ a.e., $Q(f(y|\cdot))$ is positive, continuous and bounded (by 1).

Note that in (A0) we could replace \mathcal{X} by a subset of a Polish space. Choosing $\mathcal{X} \subset \mathbb{R}^d$ is a simplification to check easily Assumptions (A0)–(A1) on the examples, especially when the hidden Markov chain comes from the discretization of a diffusion process.

We shall discuss Assumption (B3), because it is not classical. This new assumption is discussed in Section 4 and clarified on examples. It holds whenever $f(y|x)$ is uniformly lower bounded (as in [4]) but it is strictly weaker.

Some more notations are needed for the sequel. Define the family of operators: for $f : \mathcal{X} \rightarrow \mathbb{R}$ measurable and bounded, and $k \geq 0$,

$$L_k f(x) = g_k(x) Q f(x) \text{ where } g_k(x) := f(Y_k|x). \tag{2.3}$$

For $0 \leq i \leq j$, let $L_{i,j} := L_i \dots L_j$ denote the compound operator. For η a probability measure on \mathcal{X} , set

$$\Phi_{Y_k}(\eta)(f) = \frac{\eta L_k f}{\eta L_k \mathbf{1}}.$$

Then the predictive distributions satisfy $\eta_{0|-1:0}^{\eta_0} = \eta_0$ and for $k \geq 1$,

$$\eta_{k|k-1:0}^{\eta_0} f = \frac{\eta_{k-1|k-2:0}^{\eta_0} L_{k-1} f}{\eta_{k-1|k-2:0}^{\eta_0} \mathbf{1}} = \Phi_{Y_{k-1}}(\eta_{k-1|k-2:0}^{\eta_0})(f). \quad (2.4)$$

By iteration,

$$\mathbb{E}_{\eta_0}(f(X_k)|Y_{0:k-1}) = \eta_{k|k-1:0}^{\eta_0}(f) = \Phi_{Y_{k-1}} \circ \cdots \circ \Phi_{Y_0}(\eta_0)(f) = \frac{\eta_0 L_{0,k-1} f}{\eta_0 L_{0,k-1} \mathbf{1}}. \quad (2.5)$$

For δ_x the Dirac mass at x , we have

$$\eta_{k|k-1:i}^{\delta_x} f = \frac{\delta_x L_{i,k-1} f}{\delta_x L_{i,k-1} \mathbf{1}} = \Phi_{Y_{k-1}} \circ \cdots \circ \Phi_{Y_i}(\delta_x) f. \quad (2.6)$$

We will simply set $\eta_{k|k-1:i} f(x) := \eta_{k|k-1:i}^{\delta_x} f$. Moreover we have the relations

$$\mathbb{E}_{\eta_0}(f(X_k)|Y_{0:k}) = \pi_{k|k:0}^{\eta_0} f = \frac{\eta_{k|k-1:0}^{\eta_0}(g_k f)}{\eta_{k|k-1:0}^{\eta_0}(g_k)} \quad (2.7)$$

$$\text{and } \eta_{k+1|k:0}^{\eta_0}(f) = \pi_{k|k:0}^{\eta_0}(Qf). \quad (2.8)$$

Note that for all y , $\Phi_y(\delta_x)(dx') = p(x, x') dx'$. For $\eta_0(dx) = h_0(x) dx$, with h_0 positive and continuous on \mathcal{X} ,

$$\Phi_y(\eta_0)(dx') = \int_{\mathcal{X}} dx \frac{f(y|x) h_0(x) p(x, x')}{\int_{\mathcal{X}} dx f(y|x) p(x, x')} dx',$$

where the denominator is positive. Hence $\Phi_y(\eta)$ has a positive and continuous density when η is a Dirac mass or has a positive and continuous density. For these reasons and assumption (A0), all denominators appearing in our formulae are positive.

Below, for simplicity, when no confusion is possible, we omit the sub- or superscript η_0 in the distributions. Denote the number of interacting particles by N . The distribution of the bootstrap particle filter for the prediction is denoted by $\eta_{k|k-1:0}^N(f)$ and the distribution of the bootstrap particle filter for the filter is denoted by $\pi_{k|k:0}^N(f)$. The following central limit theorem is proved in [8].

Theorem 2.1. *For f a bounded measurable function and a given sequence $(Y_{0:k})$ of observations, the following convergences in distribution hold*

$$\sqrt{N}(\eta_{k|k-1:0}^N(f) - \eta_{k|k-1:0}(f)) \xrightarrow[N \rightarrow \infty]{\mathcal{L}} \mathcal{N}(0, \Delta_{k|k-1:0}(f))$$

where

$$\Delta_{k|k-1:0}(f) = \sum_{i=0}^k \frac{\eta_{i|i-1:0} \left((L_{i,k-1} (f - \eta_{k|k-1:0} f))^2 \right)}{(\eta_{i|i-1:0} L_{i,k-1} \mathbf{1})^2}, \quad (2.9)$$

and

$$\sqrt{N}(\pi_{k|k:0}^N(f) - \pi_{k|k:0}(f)) \xrightarrow[N \rightarrow \infty]{\mathcal{L}} \mathcal{N}(0, \Gamma_{k|k:0}(f))$$

where

$$\Gamma_{k|k:0}(f) = \sum_{i=0}^k \frac{\eta_{i|i-1:0} \left((L_{i,k-1} (g_k (f - \pi_{k|k:0} f)))^2 \right)}{(\eta_{k|k-1:0}(g_k))^2 (\eta_{i|i-1:0} L_{i,k-1} \mathbf{1})^2}. \quad (2.10)$$

Note that, in [8], the above theorem is proved for a wider class of functions, including functions with polynomial growth.

In the sequel, we focus on the two asymptotic variances $\Delta_{k|k-1:0}(f)$ and $\Gamma_{k|k:0}(f)$ for f bounded, when considered as functions of $Y_{0:k}$. Proposition 2.2 gives the link between the two quantities. Recall that the initial distribution is fixed equal to η_0 .

Proposition 2.2. *For f a bounded function, and k a non negative integer,*

$$\Gamma_{k|k:0}(f) = \Delta_{k|k-1:0} \left(\frac{g_k}{\eta_{k|k-1:0}(g_k)} (f - \pi_{k|k:0} f) \right) \quad (2.11)$$

and

$$\Delta_{k+1|k:0}(f) = \eta_{k+1|k:0} \left((f - \eta_{k+1|k:0} f)^2 \right) + \Gamma_{k|k:0} (Qf - \pi_{k|k:0} Qf). \quad (2.12)$$

Proof. The first formula is immediate from (2.9) and (2.10). Using (2.4)–(2.8), we get

$$\Delta_{k|k-1:0}(f) = \sum_{i=0}^k \eta_{i|i-1:0} \left(\left(\frac{L_{i,k-1} \mathbf{1}(\cdot)}{\eta_{i|i-1:0} L_{i,k-1} \mathbf{1}} \right)^2 (\eta_{k|k-1:i} f(\cdot) - \eta_{k|k-1:0} f)^2 \right) \quad (2.13)$$

$$\Gamma_{k|k:0}(f) = \sum_{i=0}^k \frac{\eta_{i|i-1:0} \left(\left(\frac{L_{i,k-1} \mathbf{1}(\cdot)}{\eta_{i|i-1:0} L_{i,k-1} \mathbf{1}} \right)^2 (\eta_{k|k-1:i}(g_k f)(\cdot) - \pi_{k|k:0} f)^2 \right)}{(\eta_{k|k-1:0}(g_k))^2}. \quad (2.14)$$

Noting that

$$\eta_{i|i-1:0} (L_{i,k-1} \mathbf{1}) \eta_{k|k-1:0} (g_k) = \eta_{i|i-1:0} (L_{i,k} \mathbf{1}),$$

we derive

$$\begin{aligned} \Delta_{k+1|k:0}(f) &= \eta_{k+1|k:0} \left((f - \eta_{k+1|k:0} f)^2 \right) \\ &\quad + \Delta_{k|k-1:0} \left(\frac{g_k}{\eta_{k|k-1:0}(g_k)} (Qf - \eta_{k+1|k:0} f) \right) \\ &= \eta_{k+1|k:0} \left((f - \eta_{k+1|k:0} f)^2 \right) + \Gamma_{k|k:0} (Qf - \eta_{k+1|k:0} f). \end{aligned} \quad \square$$

Note that [4] and [14] give these recursive formulae in a general context. We recall them in the specific case of filtering distributions, for convenience of the reader.

3. TIGHTNESS OF THE ASYMPTOTIC VARIANCES

To stress the dependence on the observations (Y_k) , we introduce another notation for $\eta_{k|k-1:0}^\nu$. For ν a probability measure, A a borelian set, $y_{0:k-1}$ a set of fixed real values, let us introduce

$$\begin{aligned} \eta_{\nu,k}[y_{0:k-1}](A) &= \frac{\mathbb{E}_\nu(\prod_{i=0}^{k-1} f(y_i|X_i) \mathbf{1}_A(X_k))}{\mathbb{E}_\nu(\prod_{i=0}^{k-1} f(y_i|X_i))} \\ &= \frac{\nu L_{0,k-1} \mathbf{1}_A}{\nu L_{0,k-1} \mathbf{1}} = \frac{\nu g_0 Q g_1 Q \dots g_{k-1} Q \mathbf{1}_A}{\nu g_0 Q g_1 Q \dots g_{k-1}} \end{aligned}$$

here, \mathbb{E}_ν denotes the expectation with respect to the distribution of the chain (X_k) with initial distribution ν and $y_{0:k-1}$ are fixed values not involved in the expectation. To ensure that all expressions are well defined, we consider probability measures ν equal to either Dirac masses or probabilities with positive continuous densities on \mathcal{X} . In the second line, we have set $g_i(x) = f(y_i|x)$ and the formula explains the backward iterations of the operators (2.3) with $Y_k = y_k$.

The following proposition proves the exponential forgetting of the initial distribution for the predictive distributions.

Proposition 3.1. *Assume (A2) and (B1) and set $\rho = 1 - \frac{\epsilon_-^2}{\epsilon_+^2}$. Then for all non negative integer k , all probability distributions ν and ν' on \mathcal{X} and all vector $y_{0:k-1}$ of real values*

$$\|\eta_{\nu,k}[y_{0:k-1}] - \eta_{\nu',k}[y_{0:k-1}]\|_{TV} \leq \rho^k,$$

where $\|\cdot\|_{TV}$ denotes the total variation distance.

Proof. The above result is generally proved for the filtering distribution (see e.g. [1], [7] and [11]). To prove it for the predictive distributions, we follow the scheme of [11]. Let $\bar{\mathcal{X}} = \mathcal{X} \times \mathcal{X}$ and denote by \bar{Q} the Markov kernel on $\bar{\mathcal{X}}$ given by

$$\bar{Q}((x, x'), A \times A') = Q(x, A)Q(x', A').$$

Set $\bar{g}_i(x, x') = g_i(x)g_i(x')$. For two probability measures ν and ν' , notice that

$$\begin{aligned} \eta_{\nu,k}[y_{0:k-1}](A) - \eta_{\nu',k}[y_{0:k-1}](A) &= \Phi_{y_{k-1}} \circ \dots \circ \Phi_{y_0}(\nu)(\mathbf{1}_A) - \Phi_{y_{k-1}} \circ \dots \circ \Phi_{y_0}(\nu')(\mathbf{1}_A) \\ &= \frac{\mathbb{E}_{\nu \otimes \nu'}(\prod_{i=0}^{k-1} \bar{g}_i(X_i, X'_i) \mathbf{1}_A(X_k)) - \mathbb{E}_{\nu' \otimes \nu}(\prod_{i=0}^{k-1} \bar{g}_i(X_i, X'_i) \mathbf{1}_A(X_k))}{\mathbb{E}_{\nu}(\prod_{i=0}^{k-1} g_i(X_i)) \mathbb{E}_{\nu'}(\prod_{i=0}^{k-1} g_i(X_i))} \end{aligned}$$

where (X_i) and (X'_i) are two independent copies of the hidden Markov chain and $\mathbb{E}_{\nu \otimes \nu'}$ denotes the expectation with respect to the distribution of the chain (X_i, X'_i) with kernel \bar{Q} and initial distribution $\nu \otimes \nu'$.

Set $\bar{\mu} = \mu \otimes \mu$, and $\bar{x} = (x, x')$. For \bar{f} a measurable non negative function, using (A2), we have

$$\epsilon_-^2 \bar{\mu}(\bar{f}) \leq \bar{Q}(\bar{x}, \bar{f}) \leq \epsilon_+^2 \bar{\mu}(\bar{f}).$$

Setting

$$\bar{Q}_0(\bar{x}, \bar{f}) = \epsilon_-^2 \bar{\mu}(\bar{f}) \quad \text{and} \quad \bar{Q}_1(\bar{x}, \bar{f}) = \bar{Q}(\bar{x}, \bar{f}) - \bar{Q}_0(\bar{x}, \bar{f}),$$

we deduce that

$$0 \leq \bar{Q}_1(\bar{x}, \bar{f}) \leq \rho \bar{Q}(\bar{x}, \bar{f}).$$

Now let us compute the numerator:

$$\begin{aligned} R_k(\nu, \nu', A) &= \mathbb{E}_{\nu \otimes \nu'} \left(\prod_{i=0}^{k-1} \bar{g}_i(X_i, X'_i) \mathbf{1}_A(X_k) \right) - \mathbb{E}_{\nu' \otimes \nu} \left(\prod_{i=0}^{k-1} \bar{g}_i(X_i, X'_i) \mathbf{1}_A(X_k) \right) \\ &= \nu \otimes \nu'(\bar{g}_0 \bar{Q} \bar{g}_1 \bar{Q} \dots \bar{g}_{k-1} \bar{Q} \mathbf{1}_{A \times \mathcal{X}}) - \nu' \otimes \nu(\bar{g}_0 \bar{Q} \bar{g}_1 \bar{Q} \dots \bar{g}_{k-1} \bar{Q} \mathbf{1}_{A \times \mathcal{X}}). \end{aligned}$$

It may be decomposed as

$$R_k(\nu, \nu', A) = \sum_{t_{0:k-1} \in \{0,1\}^k} R_k(A, t_{0:k-1})$$

where

$$\begin{aligned} R_k(A, t_{0:k-1}) &:= \nu \otimes \nu'(\bar{g}_0 \bar{Q}_{t_0} \bar{g}_1 \bar{Q}_{t_1} \dots \bar{g}_{k-1} \bar{Q}_{t_{k-1}} \mathbf{1}_{A \times \mathcal{X}}) \\ &\quad - \nu' \otimes \nu(\bar{g}_0 \bar{Q}_{t_0} \bar{g}_1 \bar{Q}_{t_1} \dots \bar{g}_{k-1} \bar{Q}_{t_{k-1}} \mathbf{1}_{A \times \mathcal{X}}). \end{aligned}$$

Assume that for an index i , $t_i = 0$. Then

$$\begin{aligned}
\nu \otimes \nu'(\bar{g}_0 \bar{Q}_{t_0} \bar{g}_1 \bar{Q}_{t_1} \dots \bar{g}_{k-1} \bar{Q}_{t_{k-1}} \mathbf{1}_{A \times \mathcal{X}}) &= \nu \otimes \nu'(\bar{g}_0 \bar{Q}_{t_0} \bar{g}_1 \bar{Q}_{t_1} \dots \bar{g}_{i-1} \bar{Q}_{t_{i-1}} \bar{g}_i) \\
&\quad \times \epsilon_-^2 \bar{\mu}(\bar{g}_{i+1} \bar{Q}_{t_{i+1}} \dots \bar{g}_{k-1} \bar{Q}_{t_{k-1}} \mathbf{1}_{A \times \mathcal{X}}) \\
&= \nu' \otimes \nu(\bar{g}_0 \bar{Q}_{t_0} \bar{g}_1 \bar{Q}_{t_1} \dots \bar{g}_{i-1} \bar{Q}_{t_{i-1}} \bar{g}_i) \\
&\quad \times \epsilon_-^2 \bar{\mu}(\bar{g}_{i+1} \bar{Q}_{t_{i+1}} \dots \bar{g}_{k-1} \bar{Q}_{t_{k-1}} \mathbf{1}_{A \times \mathcal{X}}) \\
&= \nu' \otimes \nu(\bar{g}_0 \bar{Q}_{t_0} \bar{g}_1 \bar{Q}_{t_1} \dots \bar{g}_{k-1} \bar{Q}_{t_{k-1}} \mathbf{1}_{A \times \mathcal{X}}) \\
&= \nu \otimes \nu'(\bar{g}_0 \bar{Q}_{t_0} \bar{g}_1 \bar{Q}_{t_1} \dots \bar{g}_{k-1} \bar{Q}_{t_{k-1}} \mathbf{1}_{\mathcal{X} \times A})
\end{aligned}$$

and $R_k(A, t_{0:k-1})$ vanishes except if all $t_i = 1$. Hence

$$R_k(\nu, \nu', A) = \nu \otimes \nu'(\bar{g}_0 \bar{Q}_1 \bar{g}_1 \bar{Q}_1 \dots \bar{g}_{k-1} \bar{Q}_1 (\mathbf{1}_{A \times \mathcal{X}} - \mathbf{1}_{\mathcal{X} \times A})).$$

Therefore

$$\sup_A |R_k(\nu, \nu', A)| \leq \rho^k \mathbb{E}_{\nu \otimes \nu'} \left(\prod_{i=0}^{k-1} \bar{g}_i(X_i, X'_i) \right).$$

The result follows. \square

Remark 3.2. Applying the result of Proposition 3.1 with $g_i \equiv 1$ and $\nu' = \pi$, we get $\|\nu Q^k - \pi\|_{TV} \leq \rho^k$. Thus, (A1)–(A2) imply the geometric ergodicity of (X_k) .

Let us make some comments. The result is given in [7], but the arguments of the proof are different and rely on stronger assumptions.

Proposition 3.3. *Assume (A0)–(A2), (B1)–(B2). For f a bounded measurable function and $(y_{0:k})$ a sequence of observations, the following inequalities hold*

$$\Delta_{k|k-1:0}(f) \leq \|f\|_\infty^2 \sum_{i=0}^k \eta_{i|i-1:0} \left(\left(\frac{L_{i,k-1} \mathbf{1}}{\eta_{i|i-1:0} L_{i,k-1} \mathbf{1}} \right)^2 \right) \rho^{2(k-i)}. \quad (3.1)$$

The following propositions give upper bounds for $\Delta_{k|k-1:0}(f)$.

Proof. Remark that, for all ν :

$$\eta_{\nu,k}[Y_{0:k-1}] = \Phi_{Y_{k-1}} \circ \dots \circ \Phi_{Y_i}(\eta_{\nu,i}[Y_{0:i-1}]).$$

By Proposition 3.1, we deduce, for $\nu = \eta_0$:

$$\begin{aligned}
|\eta_{k|k-1:i}(f)(x) - \eta_{k|k-1:0}^{\eta_0}(f)| &\leq \|f\|_\infty \|\eta_{\delta_x, k-i}[Y_{i:k-1}] - \eta_{\eta_0, k}(Y_{0:k-1})\|_{TV} \\
&\leq \|f\|_\infty \|\Phi_{y_{k-1}} \circ \dots \circ \Phi_{y_i}(\delta_x) - \Phi_{y_{k-1}} \circ \dots \circ \Phi_{y_i}(\eta_{\eta_0, i}[Y_{0:i-1}])\|_{TV} \\
&\leq \|f\|_\infty \rho^{k-i}.
\end{aligned}$$

Using (2.13), we get the result. \square

We stress the fact that Proposition 3.3 only relies on the exponential stability which may hold even if (A2) is not satisfied (see [11]).

Proposition 3.4. *Under Assumption (A2) and for f measurable and bounded, it holds that*

$$\Delta_{k|k-1:0}(f) \leq \|f\|_\infty^2 \frac{\epsilon_+^2}{\epsilon_-^2} \sum_{i=0}^k \eta_{i|i-1:0} \left(\left(\frac{g_i}{\eta_{i|i-1:0} g_i} \right)^2 \right) \rho^{2(k-i)}. \quad (3.2)$$

Proof. We remark that for any probability measure ν

$$\epsilon_- \nu(g_i) \mu(g_{i+1} Q \dots g_{k-1}) \leq \nu g_i Q g_{i+1} Q \dots g_{k-1} \leq \epsilon_+ \nu(g_i) \mu(g_{i+1} Q \dots g_{k-1}).$$

By (A2), since Q is Feller and the g_i 's are positive continuous, $\mu(g_{i+1} Q \dots g_{k-1})$ is positive. Applying the left inequality with $\nu = \eta_{i|i-1:0}$ and the right inequality with $\nu = \delta_x$, it comes

$$\frac{L_{i,k-1} \mathbf{1}(x)}{\eta_{i|i-1:0} L_{i,k-1} \mathbf{1}} \leq \frac{\epsilon_+}{\epsilon_-} \frac{g_i(x) \mu(g_{i+1} Q \dots g_{k-1})}{\eta_{i|i-1:0}(g_i) \mu(g_{i+1} Q \dots g_{k-1})}.$$

Thus,

$$\Delta_{k|k-1:0}(f) \leq \|f\|_\infty^2 \frac{\epsilon_+^2}{\epsilon_-^2} \sum_{i=0}^k \eta_{i|i-1:0} \left(\left(\frac{g_i}{\eta_{i|i-1:0} g_i} \right)^2 \right) \rho^{2(k-i)}. \quad \square$$

We state the main result under the additional assumption (B3). In Section 4, we show that, under (A1), (B3) is especially easy to check.

Theorem 3.5. *Assume (A0)–(A2), (B1)–(B3). Then, for all bounded function f , the sequences of variances $(\Delta_{k|k-1:0}(f))$ and $(\Gamma_{k|k:0}(f))$ are tight.*

Proof. Using that $g_i \leq 1$

$$\eta_{i|i-1:0} \left(\left(\frac{g_i}{\eta_{i|i-1:0} g_i} \right)^2 \right) \leq \frac{1}{\eta_{i|i-1:0} g_i}.$$

Setting $B_i = -\log(\eta_{i|i-1:0} g_i)$, Lemma A.1 (see the appendix) implies that the sequence $\sum_{i=0}^k e^{B_i} \rho^{2(k-i)}$ is tight with respect to k . With Proposition 3.4, we deduce that $(\Delta_{k|k-1:0}(f))_{k \geq 0}$ is tight.

Using (2.11) and $g_k \leq 1$, we obtain:

$$\Gamma_{k|k:0}(f) \leq \frac{1}{(\eta_{k|k-1:0} g_k)^2} \Delta_{k|k-1:0}(f - \pi_{k|k:0} f).$$

Since $\|f - \pi_{k|k:0} f\|_\infty \leq 2\|f\|_\infty$, the first part implies that $(\Delta_{k|k-1:0}(f - \pi_{k|k:0} f))$ is tight. By (B3), $(\eta_{k|k-1:0}(g_k))$ is also tight. The result follows. \square

We have used in the proof that $f(y|x) \leq 1$ for all $x \in \mathcal{X}$. But as we claimed previously, up to the choice of the dominating measure κ this is not a restriction.

4. DISCUSSION AND EXAMPLES

4.1. Checking of (B3)

Let us consider a hidden chain with state space $\mathcal{X} = [a, b]$ a compact interval of \mathbb{R} satisfying (A0)–(A2) (for instance a discrete sampling of a diffusion on $[a, b]$ with reflecting boundaries). Under (B2), $r(y) = \inf_{x \in \mathcal{X}} f(y|x)$ is well defined and positive. Thus, we have

$$r(Y_k) \leq \eta_{k|k-1:0}(g_k) \leq 1.$$

Therefore, the condition $\sup_{k \geq 0} \mathbf{E} |\log(r(Y_k))|^{1+\delta} < \infty$ implies (B3). In particular, when (Y_k) is stationary *i.e.* when the initial distribution of the chain is $\eta_0 = \pi$ the stationary distribution, the condition is simply $\mathbf{E} |\log(r(Y_0))|^{1+\delta} < \infty$. Let us compute $r(y)$ in some typical examples.

Example 4.1. Assume that $Y_k = X_k + \varepsilon_k$ with $\varepsilon_k \sim_{i.i.d.} \mathcal{N}(0, 1)$ and (X_k) independent of (ε_k) . The observation kernel is $F(x, dy) = \mathcal{N}(x, 1)$. Choosing the dominating measure $\kappa(dy) = \frac{1}{\sqrt{2\pi}} dy$,

$$f(y|x) = \exp\left(-\frac{(y-x)^2}{2}\right) \geq r(y)$$

where

$$|\log(r(y))| \leq \frac{1}{2} ((y-a)^2 + (y-b)^2).$$

Example 4.2. Assume that $Y_k = \sqrt{X_k} \varepsilon_k$ with $\varepsilon_k \sim_{i.i.d.} \mathcal{N}(0, 1)$, (X_k) independent of (ε_k) and $0 < a < b$. The observation kernel is $F(x, dy) = \mathcal{N}(0, x)$. Note that

$$\frac{1}{\sqrt{2\pi b}} \exp\left(-\frac{y^2}{2a}\right) \leq \frac{1}{\sqrt{2\pi x}} \exp\left(-\frac{y^2}{2x}\right) \leq \frac{1}{\sqrt{2\pi a}}.$$

Taking $\kappa(dy) = \frac{1}{\sqrt{2\pi a}} dy$, we get that

$$|\log(r(y))| \leq C + \frac{y^2}{2a}.$$

Thus, assumption (B3) is a simple moment condition on the observations which is evidently satisfied on these examples.

4.2. The case of a diffusion on a compact manifold

Consider the stochastic differential equation

$$dZ_t = b(Z_t)dt + \sigma(Z_t)dW_t$$

with a one-dimensional observation process

$$Y_{t_i} = g(Z_{t_i}) + \varepsilon_i$$

where W is a standard Brownian motion, (ε_i) is an i.i.d. sequence of $\mathcal{N}(0, 1)$ random variables, and b, σ are Lipschitz and g is smooth enough.

Assume that the diffusion process Z is valued in a compact manifold M of dimension m embedded in \mathbb{R}^d . Assume that b and σ lead to a strictly elliptic generator on M , with heat kernel $G_t(x, y)$. We refer to [1] and [5] for the following inequality

$$c_0 e^{-c_1/t} \leq G_t(x, y) \leq c_2 t^{-m/2},$$

where c_0, c_1 and c_2 are numerical constants. Assume that $t_i = i\delta$, $\delta > 0$, $i \in \mathbb{N}$, hence the observations are equally spaced in time. Then we obtain the inequality for (A2) with μ a probability distribution with positive density with respect to Lebesgue measure on M , because the transition density of the hidden Markov chain is bounded from below by a positive value. Due to the underlying diffusion process, other assumptions on the chain are verified.

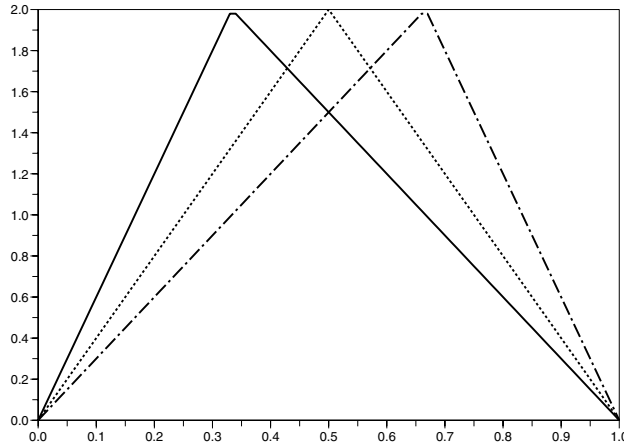


FIGURE 1. Densities involved in the toy example. Functions u (solid), v (bigdash dot) and $\frac{d\mu}{dx}$ (dot).

4.3. A toy example

Consider two continuous densities u and v on $(0, 1)$, a distribution π on $(0, 1)$ with a continuous density with respect to Lebesgue measure and a real α of $(0, 1)$. Define the Markov chain (X_k) by

$$X_0 \sim \pi, \quad X_{k+1} = \mathbf{1}_{X_k < \alpha} U_{k+1} + \mathbf{1}_{X_k \geq \alpha} V_{k+1} \quad (4.1)$$

where (U_k) and (V_k) are two independent sequences of i.i.d. random variables, independent of X_0 , with respective distributions $u(x)dx$ and $v(x)dx$. Set $p(x, x') = \mathbf{1}_{x < \alpha} u(x') + \mathbf{1}_{x \geq \alpha} v(x')$ the transition kernel density. The transition kernel admits an invariant distribution $\pi(x)dx$ with $\pi(x) = Au(x) + (1 - A)v(x)$ and

$$A = \frac{\int_0^\alpha v(x)dx}{\int_\alpha^1 u(x)dx + \int_0^\alpha v(x)dx}.$$

For example, with

$$u(x) = \begin{cases} 6x & \text{if } x \in [0, \frac{1}{3}] \\ -3x + 3 & \text{if } x \in]\frac{1}{3}, 1] \end{cases} \quad \text{and} \quad v(x) = \begin{cases} 3x & \text{if } x \in [0, \frac{2}{3}] \\ -6x + 6 & \text{if } x \in]\frac{2}{3}, 1] \end{cases}$$

the transition kernel Q of the chain (X_k) satisfies (A2) (see Fig. 1) with $\mu(dx) = 4(x \wedge 1 - x)dx$ and $\epsilon_- = \frac{1}{4}$, $\epsilon_+ = \frac{3}{2}$:

$$\forall x \in \mathcal{X}, \forall B \in \mathcal{B}(\mathcal{X}) \quad \epsilon_- \mu(B) \leq Q(x, B) \leq \epsilon_+ \mu(B).$$

In Figure 1, the graph of u is plotted in solid line, the graph of v is plotted in bigdash dotted line, and the density of μ is plotted in dotted line.

For this example, the transition density, $p(x, x')$ is not bounded from below by a positive constant. This shows that (A2) is strictly weaker than the assumption of Theorem 5 in [4]. Although Assumption (A0) is not verified on this example, the proof of the tightness still holds. Indeed, all the denominators involved in the computations of the upper bounds are well-defined and positive. Assumption (A1) is clearly verified and the stationary distribution can be explicitly computed, and the bounds in (A2) too.

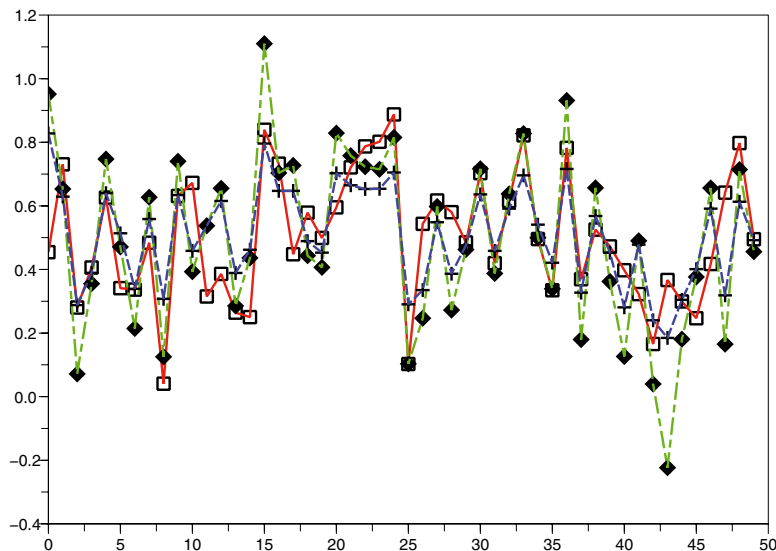


FIGURE 2. Toy example ($\alpha = 0.4$). Hidden Markov chain (plain, \square marks). Observations (longdashed line, diamond marks). Particle filter (dashed line, + marks).

TABLE 1. Computation of (5.1) for different values of N and k .

| | $\hat{\Gamma}_{50 50:0}^{(1),N}(\varphi)$ | $\hat{\Gamma}_{100 100:0}^{(1),N}(\varphi)$ | $\hat{\Gamma}_{150 150:0}^{(1),N}(\varphi)$ | $\hat{\Gamma}_{200 200:0}^{(1),N}(\varphi)$ |
|-----------|---|---|---|---|
| $N = 100$ | 6.95×10^{-2} | 4.00×10^{-2} | 4.14×10^{-2} | 5.27×10^{-2} |
| $N = 250$ | 5.91×10^{-2} | 4.01×10^{-2} | 3.35×10^{-2} | 4.27×10^{-2} |
| $N = 500$ | 6.46×10^{-2} | 3.24×10^{-2} | 3.69×10^{-2} | 4.60×10^{-2} |

5. NUMERICAL SIMULATIONS

5.1. Simulations based on the toy example

We consider Example 4.3 with the observations $Y_k = X_k + \varepsilon_k$ where $(\varepsilon_k)_k$ is a sequence of i.i.d. $\mathcal{N}(0, (0.5)^2)$ random variables. In Figure 2, we have plotted in plain line, with square marks, a trajectory of the hidden Markov chain, in longdashed line, with diamond marks, the noisy observations. We have plotted in dashed line, with plus marks, the result of the bootstrap particle filter associated to these observations, with $N = 500$ particles and $\varphi(x) = x$. We observe that the result of the particle filter is close to the hidden chain, uniformly in time.

We propose a study of the variances based on Monte-Carlo simulations.

We simulate $J = 50$ trajectories $(y_0^{(j)}, \dots, y_T^{(j)})$ for $j = 1 \dots J$ and $T = 200$. For each trajectory, we compute $L = 50$ realizations of the particular Monte-Carlo method, named $(\pi_{k|k:0}^{(j),N} \varphi)_l$ for $l = 1 \dots L$ and $k = 0 \dots T$. Hence we set

$$\hat{\Gamma}_{k|k:0}^{(j),N}(\varphi) = \frac{N}{L} \sum_{l=1}^L \left((\pi_{k|k:0}^{(j),N} \varphi)_l - \frac{1}{L} \sum_{l=1}^L (\pi_{k|k:0}^{(j),N} \varphi)_l \right)^2. \quad (5.1)$$

In Table 1, the quantity (5.1) is computed for one trajectory, and different values of the number of particles N and the number of observations k . We can notice that the value remains stable.

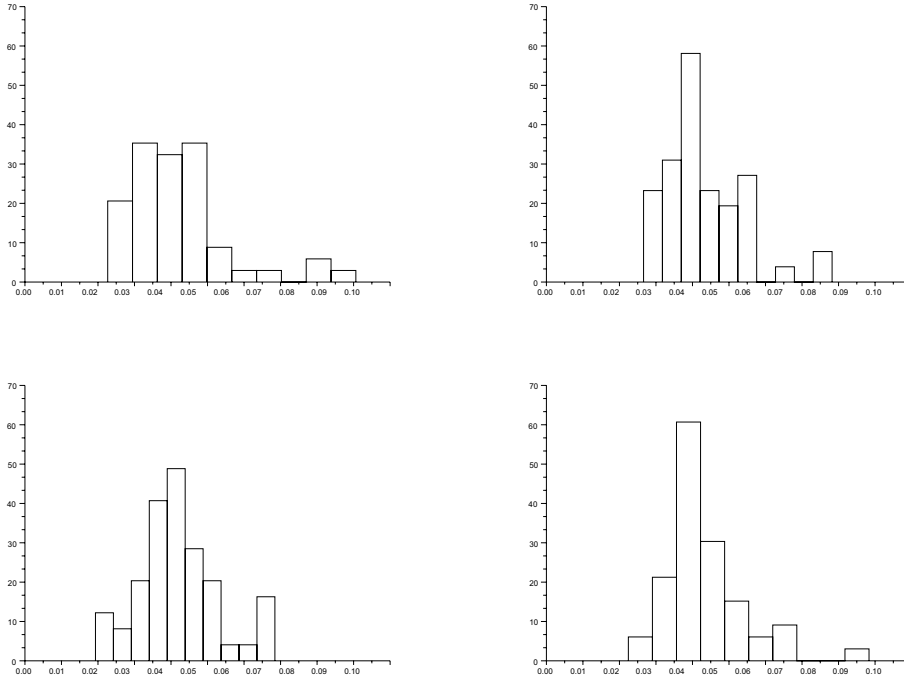


FIGURE 3. Histograms of the random variables (5.1) for $k = 50$ (top left), 100 (top right), 150 (bottom left), 200 (bottom right).

In Figure 3, the histograms of the distribution of the random variables $(\hat{\Gamma}_{k|k:0}^{(j),500}(\varphi), j = 1 \dots J)$ for $k \in \{50, 100, 150, 200\}$ are plotted. This confirms the tightness property: we do not observe any phenomenon of mass degeneracy in the distribution of the variances. Note that the small masses around the values 0.8 or 0.9 could indicate a slight degeneracy (graphs 1–2, top of Fig. 3) but appear finally under control (see graphs 3–4, bottom of Fig. 3).

Acknowledgements. The author wishes to thank Pr. Genon-Catalot, his Ph.D. advisor, for her help and the numerous discussions. He also wishes to thank Prs. Del Moral and Jacod for their advices and bibliographical references. He wishes finally to thank Mahendra Mariadassou for discussions about the numerical examples, and the anonymous referees for their suggestions and one bibliographical reference.

APPENDIX A

A.1. Bootstrap particle filter

The aim is to build a sequence of measures $(\eta_{k|k-1:0}^N)_k$, where N is the number of interacting particles, so that $\eta_{k|k-1:0}^N f$ is a good approximation of $\eta_{k|k-1:0} f$ for f bounded. We assume that the distribution of X_0 is known and we set it as $\eta_0 = \eta_{0|-1:0}$. We assume that we are able to simulate random variables under η_0 and under $Q(x, dx')$.

Step 0: Simulate $(X_0^j)_{1 \leq j \leq N}$ i.i.d. with distribution η_0 and compute $\eta_{0|-1:0}^N = \frac{1}{N} \sum_{j=1}^N \delta_{X_0^j}$.

Step 1-a: Simulate X_0^j i.i.d. with distribution $\pi_{0|0:0}^N = \sum_{j=1}^N \frac{g_0(X_0^j)}{\sum_{j=1}^N g_0(X_0^j)} \delta_{X_0^j}$.

Step 1-b: Simulate N random variables $(X_1^j)_j$ independantly with $X_1^j \sim Q(X_0^j, dx)$. Set $\eta_{1|0:0}^N = \frac{1}{N} \sum_{i=1}^N \delta_{X_1^i}$.

Step k-a: (Updating) Suppose that $\eta_{k|k-1:0}^N$ is known. Simulate $(X_k^j)_{1 \leq j \leq N}$ i.i.d. with distribution $\eta_{k|k-1:0}^N$ and simulate X_k^j i.i.d. with distribution $\pi_{k|k:0}^N = \sum_{j=1}^N \frac{g_k(X_k^j)}{\sum_{j=1}^N g_k(X_k^j)} \delta_{X_k^j}$.

Step k-b: (Prediction) Simulate X_{k+1}^N independantly with $X_{k+1}^N \sim Q(X_k^j, dx)$. Set $\eta_{k+1|k:0}^N = \frac{1}{N} \sum_{j=1}^N \delta_{X_{k+1}^j}$.

A.2. Tightness lemma

The following lemma is proved (with $\delta = 1$) in [9].

Lemma A.1. (Tightness lemma) Let $\alpha \in (0, 1)$ and consider two sequences $(A_{i,k})_{1 \leq i \leq k}$ and $(B_{i,k})_{1 \leq i \leq k}$ of non negative random variables such that

$$\sup_{i,k} \mathbb{E}(A_{i,k}) + \sup_{i,k} \mathbb{E}(B_{i,k}^{1+\delta}) = K < \infty \quad (\text{A.1})$$

then the sequence

$$\Upsilon_k = \sum_{i=1}^k \alpha^{k-i} A_{i,k} e^{B_{i,k}} \quad (\text{A.2})$$

is tight.

Proof. Choose $\gamma > 1$ such that $\alpha\gamma < 1$. Set $\Omega_{j,k} = \cap_{i=1}^{k-j} \{|B_{i,k}| \leq (k-i) \log \gamma\}$ for $1 \leq j \leq k$. Set also $\epsilon_j = \sum_{i \geq j} \frac{1}{i^{1+\delta}}$. Then, with the Markov inequality for $B_{i,k}$ we have

$$\mathbb{P}(\Omega_{j,k}^c) \leq \frac{K}{(\log \gamma)^{1+\delta}} \sum_{i=1}^{k-j} \frac{1}{(k-i)^{1+\delta}} \leq \frac{K\epsilon_j}{(\log \gamma)^{1+\delta}}.$$

On $\Omega_{j,k}$ we have

$$\begin{aligned} \sum_{i=1}^k \alpha^{k-i} A_{i,k} e^{B_{i,k}} &= \sum_{i=1}^{k-j} \alpha^{k-i} A_{i,k} e^{B_{i,k}} + \sum_{i=k-j+1}^k \alpha^{k-i} A_{i,k} e^{B_{i,k}} \\ &\leq \sum_{i=1}^{k-j} (\gamma\alpha)^{k-i} A_{i,k} + \sum_{i=k-j+1}^k \alpha^{k-i} A_{i,k} e^{B_{i,k}}. \end{aligned}$$

Finally, with the Markov inequality for $A_{i,k}$, we get for $1 \leq j \leq k$

$$\mathbb{P}(\Upsilon_k > M) \leq \frac{K\epsilon_j}{(\log \gamma)^{1+\delta}} + \frac{2K}{M(1-\alpha\gamma)} + \sum_{i=k-j+1}^k \mathbb{P}\left(A_{i,k} e^{B_{i,k}} > \frac{M}{2j}\right)$$

with our assumption, the sequence $(A_{i,k} e^{B_{i,k}})_{1 \leq i \leq k}$ is tight. For $\epsilon > 0$ we first choose j then M , hence the result. \square

REFERENCES

- [1] R. Atar and O. Zeitouni, Exponential stability for nonlinear filtering. *Ann. Inst. Henri Poincaré* **33** (1997) 697–725.
- [2] O. Cappé, E. Moulines and T. Ryden, Inference in Hidden Markov Models, in *Springer Series in Statistics*. Springer-Verlag New York, Inc., Secaucus, NJ, USA (2005).
- [3] M. Chaleyat-Maurel and V. Genon-Catalot, Computable infinite-dimensional filters with applications to discretized diffusion processes. *Stoc. Proc. Appl.* **116** (2006) 1447–1467.
- [4] N. Chopin, Central limit theorem for sequential Monte Carlo methods and its application to Bayesian inference. *Ann. Statist.* **32** (2004) 2385–2411.

- [5] E.B. Davies, Heat kernels and spectral theory, in *Cambridge Tracts in Mathematics*. Cambridge University Press, Cambridge **92** (1989).
- [6] P. Del Moral, Feynman-Kac formulae, Genealogical and interacting particle systems with applications. *Probab. Appl.* Springer-Verlag, New York (2004).
- [7] P. Del Moral and A. Guionnet, On the stability of interacting processes with applications to filtering and genetic algorithms. *Ann. Inst. Henri Poincaré* **37** (2001) 155–194.
- [8] P. Del Moral and J. Jacod, Interacting particle filtering with discrete observations, in *Sequential Monte Carlo methods in practice*, Springer, New York. *Stat. Eng. Inf. Sci.* (2001) 43–75.
- [9] P. Del Moral and J. Jacod, Interacting particle filtering with discrete-time observations: asymptotic behaviour in the Gaussian case, in *Stochastics in finite and infinite dimensions*, Birkhäuser Boston, Boston, MA. *Trends Math.* (2001) 101–122.
- [10] R. Douc, A. Guillin and J. Najim, Moderate deviations for particle filtering. *Ann. Appl. Probab.* **15** (2005) 587–614.
- [11] R. Douc, G. Fort, E. Moulines and P. Priouret, Forgetting of the initial distribution for hidden Markov models. *Stoc. Proc. Appl.* **119** (2009) 1235–1256.
- [12] A. Doucet, N. de Freitas and N. Gordon, *Sequential Monte Carlo methods in practice*, *Stat. Eng. Inform. Sci.* Springer-Verlag, New York (2001).
- [13] H.R. Künsch, State space and hidden Markov models, in *Complex Stochastic Systems*. Eindhoven (1999); Chapman & Hall/CRC, Boca Raton, FL. *Monogr. Statist. Appl. Probab.* **87** (2001) 109–173.
- [14] H.R. Künsch, Recursive Monte Carlo filters: algorithms and theoretical analysis. *Ann. Statist.* **33** (2005) 1983–2021.
- [15] N. Oudjane and S. Rubenthaler, Stability and uniform particle approximation of nonlinear filters in case of non ergodic signals. *Stoch. Anal. Appl.* **23** (2005) 421–448.
- [16] C.P. Robert and G. Casella, Monte Carlo statistical methods, 2nd edition, in *Springer Texts in Statistics*. Springer-Verlag, New York (2004).
- [17] R. Van Handel, Uniform time average consistency of Monte Carlo particle filters. *Stoc. Proc. Appl.* **119** (2009) 3835–3861.