

LIMIT THEOREMS FOR WEIGHTED SAMPLES WITH APPLICATIONS TO SEQUENTIAL MONTE CARLO METHODS

R. DOUC¹ AND E. MOULINES²

Abstract. In the last decade, sequential Monte-Carlo methods (SMC) emerged as a key tool in computational statistics (see for instance [3], [9], [7]). These algorithms approximate a sequence of distributions by a sequence of weighted empirical measures associated to a weighted population of particles, which are generated recursively.

Despite many theoretical advances (see for instance [4], [8], [2], [1]), the asymptotic property of these approximations remains a question of central interest. In this paper, we analyze sequential Monte Carlo (SMC) methods from an asymptotic perspective, that is, we establish law of large numbers and invariance principle as the number of particles gets large. We introduce the concepts of *weighted sample* consistency and asymptotic normality, and derive conditions under which the transformations of the weighted sample used in the SMC algorithm preserve these properties. To illustrate our findings, we analyze SMC algorithms to approximate the filtering distribution in state-space models. We show how our techniques allow to relax restrictive technical conditions used in previously reported works and provide grounds to analyze more sophisticated sequential sampling strategies, including branching, selection at randomly selected times, etc..

1. INTRODUCTION

Sequential Monte Carlo (SMC) refers to a class of methods designed to approximate a *sequence of probability distributions* over a *sequence of probability space* by a set of points, termed *particles* that each have an assigned non-negative weight and are updated recursively in time. SMC methods can be seen as a combination of the sequential importance sampling introduced method in [5] and the sampling importance resampling algorithm proposed in [14]; it uses a combination of *mutation* and *selection* steps. In the *mutation* step, the particles are propagated forward in time using proposal kernels and their importance weights are updated taking into account the targeted distribution. In the *selection* (or *resampling*) step, particles multiply or die depending on their *fitness* measured by their importance weights. Many algorithms have been proposed since, which differ in the way the particles and the importance weights evolve and adapt.

In this paper, we focus on the asymptotic behavior of the *weighted particle* approximation as the number of particles tend to infinity. Because the particles interact during the selection steps, they are not independent which make the analysis of particle approximation a challenging area of research. The main purpose of this paper is to derive an asymptotic theory of weighted system of particles. To our best knowledge, limit theorems for such weighted approximations were only considered in [11], who mostly sketched consistency proofs. In this paper, we establish both law of large numbers and central limit theorems, under assumptions that are presumably closed from being minimal. These results apply not only to the many different implementations of the SMC algorithms. They cover resampling schedules (when to resample) that can be either deterministic or dynamic, *i.e.* based on the distribution of the importance weights at the current iteration. We do not impose

¹ CMAP, École Polytechnique, Palaiseau, France

² Laboratoire Traitement et Communication de l'Information, CNRS / GET Télécom Paris, France

a specific structure on the sequence of the target probability measure; therefore our results apply not only to sequential filtering or smoothing of state-space contexts, but also to recent algorithms developed for population Monte Carlo or for molecular simulation.

2. DEFINITIONS AND MAIN RESULTS

Let Ξ be a general state space, μ be a probability measure on $(\Xi, \mathcal{B}(\Xi))$, $\{M_N\}_{N \geq 0}$ be a sequence of integers, and \mathcal{C} be a subset of $L^1(\Xi, \mu)$. We approximate the probability measure μ by points $\xi_{N,i} \in \Xi$, $i = 1, \dots, M_N$ associated to non-negative weights $\omega_{N,i} \geq 0$.

Definition 1. A weighted sample $\{(\xi_{N,i}, \omega_{N,i})\}_{1 \leq i \leq M_N}$ on Ξ is said to be consistent for the probability measure μ and the set $\mathcal{C} \subseteq L^1(\Xi, \mu)$ if for any $f \in \mathcal{C}$, as $N \rightarrow \infty$,

$$\Omega_N^{-1} \sum_{i=1}^{M_N} \omega_{N,i} f(\xi_{N,i}) \xrightarrow{P} \mu(f), \quad (1)$$

$$\Omega_N^{-1} \max_{1 \leq i \leq M_N} \omega_{N,i} \xrightarrow{P} 0 \quad \text{where} \quad \Omega_N = \sum_{i=1}^{M_N} \omega_{N,i}. \quad (2)$$

This definition of weighted sample consistency is similar to the notion of *properly weighted sample* introduced in [11]. The difference stems from the condition (2) which implies that the contribution of each individual term in the sum vanishes in the limit as $N \rightarrow \infty$, a condition often referred as *smallness* in the literature.

In order to obtain sensible results we restrict our attention to sets of functions which are sufficiently rich.

Definition 2 (Proper Set). A set \mathcal{C} of real-valued measurable functions on Ξ is said to be proper if the following conditions are satisfied.

- (i) \mathcal{C} is a linear space: for any f and g in \mathcal{C} and reals α and β , $\alpha f + \beta g \in \mathcal{C}$,
- (ii) If $g \in \mathcal{C}$ and f is measurable with $|f| \leq |g|$, then $f \in \mathcal{C}$,
- (iii) For all c , the constant function $f \equiv c$ belongs to \mathcal{C} .

For any function f , define the positive and negative parts of it by

$$f^+ \stackrel{\text{def}}{=} f \vee 0 \quad \text{and} \quad f^- \stackrel{\text{def}}{=} (-f) \vee 0,$$

and note that f^+ and f^- are both dominated by $|f|$. Thus, if $|f| \in \mathcal{C}$ then f^+ and f^- both belong to \mathcal{C} and so does $f = f^+ - f^-$. It is easily seen that for any $p \geq 0$ and any measure μ on $(\Xi, \mathcal{B}(\Xi))$, the set $L^p(\Xi, \mu)$ is proper.

Let μ be a probability measure on $(\Xi, \mathcal{B}(\Xi))$, γ be a finite measure on $(\Xi, \mathcal{B}(\Xi))$, $\mathcal{A} \subseteq L^1(\Xi, \mu)$ and $\mathcal{W} \subseteq L^1(\Xi, \gamma)$ be sets of real-valued measurable functions on Ξ , σ a real non negative function on \mathcal{A} , and $\{a_N\}$ be a non-decreasing real sequence diverging to infinity.

Definition 3. A weighted sample $\{(\xi_{N,i}, \omega_{N,i})\}_{1 \leq i \leq M_N}$ on Ξ is said to be asymptotically normal for $(\mu, \mathcal{A}, \mathcal{W}, \sigma, \gamma, \{a_N\})$ if

$$a_N \Omega_N^{-1} \sum_{i=1}^{M_N} \omega_{N,i} \{f(\xi_{N,i}) - \mu(f)\} \xrightarrow{\mathcal{D}} \mathcal{N}\{0, \sigma^2(f)\} \quad \text{for any } f \in \mathcal{A}, \quad (3)$$

$$a_N^2 \Omega_N^{-2} \sum_{i=1}^{M_N} \omega_{N,i}^2 f(\xi_{N,i}) \xrightarrow{P} \gamma(f) \quad \text{for any } f \in \mathcal{W} \quad (4)$$

$$a_N \Omega_N^{-1} \max_{1 \leq i \leq M_N} \omega_{N,i} \xrightarrow{P} 0 \quad (5)$$

To analyze the sequential Monte Carlo methods discussed in the introduction, we now need to study how the mutation step and the resampling step affects the consistent or / and asymptotically normal weighted sample.

2.1. Mutation

To study SISR algorithms, we need first to show that when moving the particles using a Markov kernel and then assigning them appropriately defined importance weights, we transform a weighted sample consistent (or asymptotically normal) for a distribution ν on a general state space $(\Xi, \mathcal{B}(\Xi))$ into a weighted sample consistent (or asymptotically normal) for a distribution μ on $(\tilde{\Xi}, \mathcal{B}(\tilde{\Xi}))$. Let L be a kernel from $(\Xi, \mathcal{B}(\Xi))$ to $(\tilde{\Xi}, \mathcal{B}(\tilde{\Xi}))$ such that $\nu L(\tilde{\Xi}) > 0$ and for any $f \in \mathbb{B}(\tilde{\Xi})$,

$$\mu(f) = \frac{\nu L(f)}{\nu L(\tilde{\Xi})} = \frac{\iint \nu(d\xi) L(\xi, d\tilde{\xi}) f(\tilde{\xi})}{\iint \nu(d\xi) L(\xi, \tilde{\Xi})} . \tag{6}$$

There exist of course many such kernels: one may set for example $L(\xi, \tilde{A}) = \mu(\tilde{A})$ for any $\xi \in \Xi$, but, as we will see below, this is not usually the most appropriate choice.

We wish to transform a weighted sample $\{(\xi_{N,i}, \omega_{N,i})\}_{1 \leq i \leq M_N}$ targeting the distribution ν on $(\Xi, \mathcal{B}(\Xi))$ into a weighted sample $\{(\tilde{\xi}_{N,i}, \tilde{\omega}_{N,i})\}_{1 \leq i \leq \tilde{M}_N}$ targeting μ on $(\tilde{\Xi}, \mathcal{B}(\tilde{\Xi}))$. The number of particles \tilde{M}_N is set to be $\tilde{M}_N = \alpha_N M_N$ where α_N is the number of offsprings of each particle. These offsprings are proposed using Markov kernels $R, k = 1, \dots, \alpha_N$, from $(\Xi, \mathcal{B}(\Xi))$ to $(\tilde{\Xi}, \mathcal{B}(\tilde{\Xi}))$.

Most importantly, we assume that for any $\xi \in \Xi$, the probability measure $L(\xi, \cdot)$ on $(\tilde{\Xi}, \mathcal{B}(\tilde{\Xi}))$ is absolutely continuous with respect to R , which we denote $L(\xi, \cdot) \ll R(\xi, \cdot)$ and define

$$W(\xi, \tilde{\xi}) \stackrel{\text{def}}{=} \frac{dL(\xi, \cdot)}{dR(\xi, \cdot)}(\tilde{\xi}) \tag{7}$$

We draw new particle positions $\{\tilde{\xi}_{N,j}\}_{1 \leq j \leq \tilde{M}_N}$ conditionally independently given $\mathcal{F}_{N,0} = \sigma(\{(\xi_{N,i}, \omega_{N,i})\}_{1 \leq i \leq M_N})$ with distribution given for $i = 1, \dots, M_N, k = 1, \dots, \alpha_N$ and $A \in \mathcal{B}(\tilde{\Xi})$ by

$$P\left(\tilde{\xi}_{N, \alpha_N(i-1)+k} \in A \mid \mathcal{F}_{N,0}\right) = R(\xi_{N,i}, A) \tag{8}$$

and associate to each new particle positions the importance weight:

$$\tilde{\omega}_{N, \alpha_N(i-1)+k} = \omega_{N,i} W(\xi_{N,i}, \tilde{\xi}_{N, \alpha_N(i-1)+k}), \text{ where } i = 1, \dots, M_N, k = 1, \dots, \alpha_N . \tag{9}$$

The mutation step is *unbiased* in the sense that, for any $f \in \mathcal{B}(\tilde{\Xi})$ and $i = 1, \dots, M_N$,

$$\sum_{j=\alpha_N(i-1)+1}^{\alpha_N i} E\left[\tilde{\omega}_{N,j} f(\tilde{\xi}_{N,j}) \mid \mathcal{F}_{N,j-1}\right] = \alpha_N \omega_{N,i} L(\xi_{N,i}, f), \tag{10}$$

where for $j = 1, \dots, \tilde{M}_N, \mathcal{F}_{N,j} \stackrel{\text{def}}{=} \mathcal{F}_{N,0} \vee \sigma(\{\tilde{\xi}_{N,l}\}_{1 \leq l \leq j})$. The following theorems state conditions under which the mutation step described above preserves the weighted sample consistency.

Theorem 1. *Let ν be a probability measure on $(\Xi, \mathcal{B}(\Xi))$, μ be a probability measure on $(\tilde{\Xi}, \mathcal{B}(\tilde{\Xi}))$. Let L be a finite kernel from $(\Xi, \mathcal{B}(\Xi))$ to $(\tilde{\Xi}, \mathcal{B}(\tilde{\Xi}))$ satisfying $\nu L(\tilde{\Xi}) > 0$ any $\tilde{A} \in \mathcal{B}(\tilde{\Xi}), \mu(\tilde{A}) = \nu L(\tilde{A}) / \nu L(\tilde{\Xi})$. Assume that*

- (i) *the weighted sample $\{(\xi_{N,i}, \omega_{N,i})\}_{1 \leq i \leq M_N}$ is consistent for (ν, C) , where $C \subset \mathbb{L}^1(\Xi, \nu)$ is a proper set ,*
- (ii) *the function $L(\cdot, \tilde{\Xi})$ belongs to C ,*

(iii) for any $\xi \in \Xi$, $L(\xi, \cdot) \ll R(\xi, \cdot)$.

Then, $\tilde{\mathcal{C}}$ is a proper set and the weighted sample $\{(\tilde{\xi}_{N,i}, \tilde{\omega}_{N,i})\}_{1 \leq i \leq \tilde{M}_N}$ defined by (8) and (9) is consistent for $(\mu, \tilde{\mathcal{C}})$, where

$$\tilde{\mathcal{C}} \stackrel{\text{def}}{=} \{f \in L^1(\tilde{\Xi}, \mu), L(\cdot, |f|) \in \mathcal{C}\} , \quad (11)$$

We now turn to prove the asymptotic normality.

Theorem 2. Suppose that the assumptions of Theorem 1 hold. Let γ be a finite measure on $(\Xi, \mathcal{B}(\Xi))$, $A \subseteq L^1(\Xi, \nu)$, and $W \subseteq L^1(\Xi, \gamma)$ be proper sets, σ be a non negative function on A . Assume in addition that

- (i) the weighted sample $\{(\xi_{N,i}, \omega_{N,i})\}_{1 \leq i \leq M_N}$ is asymptotically normal for $(\nu, A, W, \sigma, \gamma, \{M_N^{1/2}\})$,
- (ii) the function $R(\cdot, W^2)$ belongs to W .

Then, if $\alpha_N \rightarrow \alpha$, the weighted sample $\{(\tilde{\xi}_{N,i}, \tilde{\omega}_{N,i})\}_{1 \leq i \leq \tilde{M}_N}$ is asymptotically normal for $(\mu, \tilde{A}, \tilde{W}, \tilde{\sigma}, \tilde{\gamma}, \{M_N^{1/2}\})$ with

$$\begin{aligned} \tilde{A} &\stackrel{\text{def}}{=} \left\{ f \in L^1(\tilde{\Xi}, \mu) : L(\cdot, |f|) \in A \quad \text{and} \quad R(\cdot, W^2 f^2) \in W \right\} , \\ \tilde{W} &\stackrel{\text{def}}{=} \left\{ f \in L^1(\tilde{\Xi}, \mu) : R(\cdot, W^2 |f|) \in W \right\} , \\ \tilde{\sigma}^2(f) &\stackrel{\text{def}}{=} \sigma^2\{\tilde{L}[f - \mu(f)]\} + \alpha^{-1} \gamma R\{[\tilde{W}f - R(\cdot, \tilde{W}f)]^2\} , & \text{for all } f \in \tilde{A} , \\ \tilde{\gamma}(f) &\stackrel{\text{def}}{=} \alpha^{-1} \gamma R(\tilde{W}^2 f) , & \text{for all } f \in \tilde{W} , \end{aligned}$$

where $\tilde{L} \stackrel{\text{def}}{=} [\nu L(\tilde{\Xi})]^{-1} L$ and $\tilde{W} = [\nu L(\tilde{\Xi})]^{-1} W$. Moreover, \tilde{A} and \tilde{W} are proper sets.

Note that reweighting the particles without moving them is a particular case of mutation. Thus, Theorem 1 and 2 may also apply in this context.

2.2. Resampling

Resampling is the second basic type of transformation used in sequential Monte-Carlo methods; resampling converts a weighted sample $\{(\xi_{N,i}, \omega_{N,i})\}_{1 \leq i \leq M_N}$ targeting a distribution ν into an equally weighted sample $\{(\tilde{\xi}_{N,i}, 1)\}_{1 \leq i \leq \tilde{M}_N}$ targeting the same distribution ν on $(\Xi, \mathcal{B}(\Xi))$. We will focus on importance sampling estimators that satisfy the following *unbiasedness condition*:

$$\text{for any } f \in \mathbb{B}(\Xi), \quad \mathbb{E} \left[\tilde{M}_N^{-1} \sum_{i=1}^{\tilde{M}_N} f(\tilde{\xi}_{N,i}) \middle| \mathcal{F}_N \right] = \Omega_N^{-1} \sum_{i=1}^{M_N} \omega_{N,i} f(\xi_{N,i}) , \quad (12)$$

where $\mathcal{F}_N = \sigma(\{(\xi_{N,i}, \omega_{N,i})\}_{1 \leq i \leq M_N})$. There are many different unbiased resampling procedures described in the literature. The most elementary procedure is the so-called *multinomial* sampling in which we draw conditionally independently given \mathcal{F}_N integer-valued random variables $\{I_{N,k}\}_{1 \leq k \leq \tilde{M}_N}$ with distribution

$$\mathbb{P}(I_{N,k} = i | \mathcal{F}_N) = \Omega_N^{-1} \omega_{N,i} , \quad i = 1, \dots, M_N , \quad (13)$$

and set

$$\tilde{\xi}_{N,i} = \xi_{N, I_{N,i}} , \quad \text{for } i = 1, \dots, \tilde{M}_N . \quad (14)$$

If the weighted sample $\{(\xi_{N,i}, \omega_{N,i})\}_{1 \leq i \leq M_N}$ is consistent for (ν, \mathcal{C}) , where \mathcal{C} is a proper subset of $\mathbb{B}(X)$, it is a natural question to ask whether $\{(\tilde{\xi}_{N,i}, 1)\}_{1 \leq i \leq \tilde{M}_N}$ is consistent for ν and, if so, what an appropriately defined class of functions on Ξ might be. It happens that a fairly general result can be obtained in this case.

Theorem 3. Let ν a probability measure on $(\Xi, \mathcal{B}(\Xi))$ and $C \subseteq L^1(\Xi, \nu)$ be a proper set of functions. Assume that the weighted sample $\{(\xi_{N,i}, \omega_{N,i})\}_{1 \leq i \leq M_N}$ is consistent for (ν, C) , where $C \subseteq \mathbb{B}(\Xi)$ is a proper set. Then, the uniformly weighted sample $\{(\tilde{\xi}_{N,i}, 1)\}_{1 \leq i \leq \tilde{M}_N}$ obtained using **any** unbiased resampling scheme (i.e. satisfying (12)) is consistent for (ν, C) .

It is also sensible in this discussion to strengthen the requirement of consistency into asymptotic normality, and again prove that the sampling operation transform an asymptotically normal weighted sample for ν into an asymptotically normal sample for ν (for appropriately defined class of functions, normalizing factors, etc.). We consider first the multinomial sampling algorithm.

Theorem 4. Suppose that the assumptions of Theorem 3 hold. Let γ be a finite measure on $(\Xi, \mathcal{B}(\Xi))$, $A \subseteq L^1(\Xi, \nu)$ and $W \subseteq L^1(\Xi, \gamma)$ be proper sets, σ be a non negative function on A and $\{a_N\}$ be a non-negative sequence. Define

$$\tilde{A} \stackrel{\text{def}}{=} \{f \in A, f^2 \in C\}, \tag{15}$$

Assume in addition that

- (i) the weighted sample $\{(\xi_{N,i}, \omega_{N,i})\}_{1 \leq i \leq M_N}$ is asymptotically normal for $(\nu, A, W, \sigma, \gamma, \{a_N\})$
- (ii) $\lim_{N \rightarrow \infty} a_N = \infty$ and $\lim_{N \rightarrow \infty} a_N^2 / M_N = \beta$, where $\beta \in [0, \infty]$.

Then \tilde{A} is a proper set and the following holds true for the resampled system $\{(\tilde{\xi}_{N,i}, 1)\}_{1 \leq i \leq \tilde{M}_N}$ defined as in (13) and (14).

- (i) If $\beta < 1$, then $\{(\tilde{\xi}_{N,i}, 1)\}_{1 \leq i \leq \tilde{M}_N}$ is asymptotically normal for $(\nu, \tilde{A}, C, \tilde{\sigma}, \tilde{\gamma}, \{a_N\})$ with

$$\begin{aligned} \tilde{\sigma}^2(f) &= \beta \text{Var}_\nu(f) + \sigma^2(f) \quad \text{for any } f \in \tilde{A}, \\ \tilde{\gamma} &= \beta \nu. \end{aligned}$$

- (ii) If $\beta \geq 1$, then $\{(\tilde{\xi}_{N,i}, 1)\}_{1 \leq i \leq \tilde{M}_N}$ is asymptotically normal for $(\nu, \tilde{A}, C, \tilde{\sigma}, \tilde{\gamma}, \{\tilde{M}_N^{1/2}\})$ with

$$\begin{aligned} \tilde{\sigma}^2(f) &= \text{Var}_\nu(f) + \beta^{-1} \sigma^2(f) \quad \text{for any } f \in \tilde{A}, \\ \tilde{\gamma} &= \nu. \end{aligned}$$

3. AN APPLICATION TO STATE-SPACE MODELS

In this section, we apply the results developed above to state-space models. The state process is a Markov chain $\{X_k\}_{k \geq 1}$ is a Markov chain on a general state space X with initial distribution χ and kernel Q . The observations $\{Y_k\}_{k \geq 1}$ are random variables taking value in Y that are independent conditionally on the state sequence $\{X_k\}_{k \geq 1}$; in addition, there exists a measure λ on $(Y, \mathcal{B}(Y))$, and a transition density function $x \mapsto g(x, y)$, referred to as the likelihood, such that $P(Y_k \in A | X_k = x) = \int_A g(x, y) \lambda(dy)$, for all $A \in Y$. The kernel Q and the likelihood functions $x \mapsto g(x, y)$ are assumed to be known.

In this paper, we are primarily concerned with the recursive estimation of the (joint) smoothing distribution $\phi_{\chi, k}(y_{1:k}, \cdot)$ defined as

$$\phi_{\chi, k}(y_{1:k}, f) \stackrel{\text{def}}{=} E[f(X_{1:k}) | Y_{1:k} = y_{1:k}], \tag{16}$$

where $f \in \mathbb{B}(X^k)$ and for any sequence $\{a_i\}_{1 \leq i \leq k}$, $a_{1:k} \stackrel{\text{def}}{=} (a_1, \dots, a_k)$. We shall consider the case in which the observations have an arbitrary but fixed value $y_{1:k}$, and we drop them from the notations. We denote $g_k(x) = g(x, y_k)$. In the Monte-Carlo framework, we approximate the posterior distribution $\phi_{\chi, k}$ at each iteration k by means of a weighted sample $\{(\xi_{N,i}^{(k)}, \omega_{N,i}^{(k)})\}_{1 \leq i \leq M_N}$. Note that $\phi_{\chi, k}$ is defined on $(X^k, \mathcal{B}(X^k))$ and thus that the points $\xi_{N,i}^{(k)}$ belong to X^k (these are often referred to as *path particle* in the literature; see [2]).

We will apply the results presented in section 2, it is first required to define a transition kernel L_{k-1} satisfying (6) with $\nu = \phi_{\chi, k-1}$, $(\Xi, \mathcal{B}(\Xi)) = (X^{k-1}, \mathcal{B}(X^{k-1}))$, $\mu = \phi_{\chi, k}$ and $(\tilde{\Xi}, \mathcal{B}(\tilde{\Xi})) = (X^k, \mathcal{B}(X^k))$, i.e. for any $f \in$

$\mathbb{B}^+(\mathsf{X}^k)$,

$$\phi_{\chi,k}(f) = \frac{\int \cdots \int \phi_{\chi,k-1}(dx_{1:k-1}) L_{k-1}(x_{1:k-1}, d\tilde{x}_{1:k}) f(\tilde{x}_{1:k})}{\int \cdots \int \phi_{\chi,k-1}(dx_{1:k-1}) L_{k-1}(dx_{1:k-1}, \mathsf{X}^k)} \quad (17)$$

Then, we must choose a proposal kernel R_{k-1} satisfying

$$L_{k-1}(x_{1:k-1}, \cdot) \ll R_{k-1}(x_{1:k-1}, \cdot), \quad \text{for any } x_{1:k-1} \in \mathsf{X}^{k-1}. \quad (18)$$

There are many possible choices, which will be associated with different algorithms proposed in the literature. The first obvious choice consists in setting, for any $f \in \mathcal{B}(\mathsf{X}^k)$,

$$L_{k-1}(x_{1:k-1}, f) = T_{k-1}(x_{1:k-1}, f) = \int Q(x_{k-1}, d\tilde{x}_k) g_k(\tilde{x}_k) f(x_{1:k-1}, \tilde{x}_k). \quad (19)$$

Note that, by construction, the kernel $T_{k-1}(x_{1:k-1}, \cdot)$ leaves the coordinates $x_{1:k-1}$ unchanged.

The algorithm goes as follows. We proceed from $\{(\xi_{N,i}^{(k-1)}, \omega_{N,i}^{(k-1)})\}_{1 \leq i \leq M_N}$ targeting $\phi_{\chi,k-1}$ to $\{(\xi_{N,i}^{(k)}, \omega_{N,i}^{(k)})\}_{1 \leq i \leq M_N}$ targeting $\phi_{\chi,k}$ as follows. To keep the discussion simple, it is assumed that each particle gives birth to a single offspring. In the mutation step, we draw $\{\tilde{\xi}_{N,i}^{(k)}\}_{1 \leq i \leq M_N}$ conditionally independently given $\mathcal{F}_N^{(k-1)}$ with distribution given, for any $f \in \mathbb{B}^+(\mathsf{X}^k)$ by

$$\mathbb{E} \left[f(\tilde{\xi}_{N,i}^{(k)}) \mid \mathcal{F}_N^{(k-1)} \right] = R_{k-1}(\xi_{N,i}^{(k-1)}, f) = \int \cdots \int R_{k-1}(\xi_{N,i}^{(k-1)}, dx_{1:k}) f(x_{1:k}), \quad (20)$$

where $i = 1, \dots, M_N$. Next we assign to the particle $\tilde{\xi}_{N,i}^{(k)}$, $i = 1, \dots, M_N$, the importance weight

$$\tilde{\omega}_{N,i}^{(k)} = \omega_{N,i}^{(k-1)} W_{k-1}(\xi_{N,i}^{(k-1)}, \tilde{\xi}_{N,i}^{(k)}) \quad \text{with} \quad W_{k-1}(x_{1:k-1}, \tilde{x}_{1:k}) = \frac{dL_{k-1}(x_{1:k-1}, \cdot)}{dR_{k-1}(x_{1:k-1}, \cdot)}(\tilde{x}_{1:k}). \quad (21)$$

Instead of resampling at each iteration (which is the assumption upon which most of the asymptotic analysis have been carried out so far), we rejuvenate the particle system only when the importance weights are too skewed. As discussed in [6, section 4], a sensible approach is to try to monitor the coefficient of variations of weight, defined by

$$[\text{CV}_N^{(k)}]^2 \stackrel{\text{def}}{=} \frac{1}{M_N} \sum_{i=1}^{M_N} \left(\frac{M_N \tilde{\omega}_{N,i}^{(k)}}{\tilde{\Omega}_N^{(k)}} - 1 \right)^2.$$

The coefficient of variation is minimal when the normalized importance weights $\tilde{\omega}_{N,i}^{(k)}/\tilde{\Omega}_N^{(k)}$, $i = 1, \dots, M_N$, are all equal to $1/M_N$, in which case $\text{CV}_N^{(k)} = 0$. The maximal value of $\text{CV}_N^{(k)}$ is $\sqrt{M_N - 1}$, which corresponds to one of the normalized weights being one and all others being null. If the coefficient of variation of the importance weights crosses a threshold we rejuvenate the particle system. More precisely if at time k , $\text{CV}_N^{(k)} \geq \kappa$, we draw $I_k^{N,1}, \dots, I_k^{N,M_N}$ conditionally independently given $\tilde{\mathcal{F}}_N^{(k)} = \mathcal{F}_N^{(k-1)} \vee \sigma(\{\tilde{\xi}_{N,i}^{(k)}, \tilde{\omega}_{N,i}^{(k)}\}_{1 \leq i \leq M_N})$, with distribution

$$\mathbb{P} \left(I_k^{N,i} = j \mid \tilde{\mathcal{F}}_N^{(k)} \right) = \tilde{\omega}_{N,j}^{(k)} / \tilde{\Omega}_N^{(k)}, \quad i = 1, \dots, M_N, j = 1, \dots, M_N \quad (22)$$

and we set

$$\xi_{N,i}^{(k)} = \tilde{\xi}_{N,I_k^{N,i}}^{(k)} \quad \text{and} \quad \omega_{N,i}^{(k)} = 1, \quad i = 1, \dots, M_N. \quad (23)$$

If $\text{CV}_N^{(k)} < \kappa$, we simply copy the mutated path particles: $(\xi_{N,i}^{(k)}, \omega_{N,i}^{(k)}) = (\tilde{\xi}_{N,i}^{(k)}, \tilde{\omega}_{N,i}^{(k)})$, $i = 1, \dots, M_N$. In both cases, we set $\mathcal{F}_N^{(k)} = \tilde{\mathcal{F}}_N^{(k)} \vee \sigma(\{(\xi_{N,i}^{(k)}, \omega_{N,i}^{(k)})\}_{1 \leq i \leq M_N})$. We consider here only multinomial resampling, but the deterministic plus residual sampling can be applied as well.

Theorem 5. For any $k > 0$, let L_k and R_k be transition kernels from $(\mathbf{X}^k, \mathcal{B}(\mathbf{X}^k))$ to $(\mathbf{X}^{k+1}, \mathcal{B}(\mathbf{X}^{k+1}))$ satisfying (17) and (18), respectively. Assume that the equally weighted sample $\{(\xi_{N,i}^{(1)}, 1)\}_{1 \leq i \leq M_N}$ is consistent for $\{\phi_{\chi,1}, \mathbb{L}^1(\mathbf{X}, \phi_{\chi,1})\}$ and asymptotically normal for $(\phi_{\chi,1}, \mathbf{A}_1, \mathbf{W}_1, \sigma_1, \phi_{\chi,1}, \{M_N^{1/2}\})$ where \mathbf{A}_1 and \mathbf{W}_1 are proper sets and define recursively (\mathbf{A}_k) and (\mathbf{W}_k) by

$$\begin{aligned} \mathbf{A}_k &\stackrel{\text{def}}{=} \{f \in \mathbb{L}^2(\mathbf{X}^k, \phi_{\chi,k}), \quad L_{k-1}(\cdot, f) \in \mathbf{A}_{k-1}, R_{k-1}(\cdot, W_{k-1}^2 f^2) \in \mathbf{W}_{k-1}\}, \\ \mathbf{W}_k &\stackrel{\text{def}}{=} \{f \in \mathbb{L}^1(\mathbf{X}^k, \phi_{\chi,k}), \quad R_{k-1}(\cdot, W_{k-1}^2 |f|) \in \mathbf{W}_{k-1}\}. \end{aligned}$$

Assume in addition that for any $k \geq 1$, $R_k(\cdot, W_k^2) \in \mathbf{W}_k$. Then for any $k \geq 1$, (\mathbf{A}_k) and (\mathbf{W}_k) are proper sets and $\{(\xi_{N,i}^{(k)}, \omega_{N,i}^{(k)})\}_{1 \leq i \leq M_N}$ is consistent for $\{\phi_{\chi,k}, \mathbb{L}^1(\mathbf{X}, \phi_{\chi,k})\}$ and asymptotically normal for $(\phi_{\chi,k}, \mathbf{A}_k, \mathbf{W}_k, \sigma_k, \gamma_k, \{M_N^{1/2}\})$, where the functions σ_k and the measure γ_k are given by

$$\begin{aligned} \sigma_k^2(f) &= \varepsilon_k \text{Var}_{\phi_{\chi,k}}(f) \\ &+ \frac{\sigma_{k-1}^2(L_{k-1}\{f - \phi_{\chi,k}(f)\}) + \gamma_{k-1} R_{k-1} \left[\{W_{k-1}f - R_{k-1}(\cdot, W_{k-1}f)\}^2 \right]}{\{\phi_{\chi,k-1} L_{k-1}(\mathbf{X}^k)\}^2} \\ \gamma_k(f) &= \varepsilon_k \phi_{\chi,k} + (1 - \varepsilon_k) \frac{\gamma_{k-1} R_{k-1}(W_{k-1}^2 f)}{[\phi_{\chi,k-1} L_{k-1}(\mathbf{X}^k)]^2} \end{aligned}$$

where W_k is defined in (21) and

$$\varepsilon_k \stackrel{\text{def}}{=} \mathbf{1} \left\{ [\phi_{\chi,k-1} L_{k-1}(\mathbf{X}^k)]^{-2} \gamma_{k-1} R_{k-1}(W_{k-1}^2) \geq 1 + \kappa^2 \right\}$$

REFERENCES

- [1] N. Chopin. Central limit theorem for sequential monte carlo methods and its application to bayesian inference. *Ann. Statist.*, 32(6):2385–2411, 2004.
- [2] P. Del Moral. *Feynman-Kac Formulae. Genealogical and Interacting Particle Systems with Applications*. Springer, 2004.
- [3] A. Doucet, N. De Freitas, and N. Gordon, editors. *Sequential Monte Carlo Methods in Practice*. Springer, New York, 2001.
- [4] W. R. Gilks and C. Berzuini. Following a moving target—Monte Carlo inference for dynamic Bayesian models. *J. Roy. Statist. Soc. Ser. B*, 63(1):127–146, 2001.
- [5] J. Handschin and D. Mayne. Monte Carlo techniques to estimate the conditionnal expectation in multi-stage non-linear filtering. In *Int. J. Control*, volume 9, pages 547–559, 1969.
- [6] A. Kong, J. S. Liu, and W. Wong. Sequential imputation and Bayesian missing data problems. *J. Am. Statist. Assoc.*, 89(278–288):590–599, 1994.
- [7] H. R. Künsch. State space and hidden markov models. In O. E. Barndorff-Nielsen, D. R. Cox, and C. Klueppelberg, editors, *Complex Stochastic Systems*, pages 109–173. CRC Publisher, Boca raton, 2001.
- [8] H. R. Künsch. Recursive Monte-Carlo filters: algorithms and theoretical analysis. *Ann. Statist.*, 33(5):1983–2021, 2005.
- [9] J. Liu. *Monte Carlo Strategies in Scientific Computing*. Springer, New York, 2001.
- [10] J. Liu and R. Chen. Blind deconvolution via sequential imputations. *J. Roy. Statist. Soc. Ser. B*, 430:567–576, 1995.
- [11] J. Liu and R. Chen. Sequential Monte-Carlo methods for dynamic systems. *J. Roy. Statist. Soc. Ser. B*, 93:1032–1044, 1998.
- [12] J. Liu, R. Chen, and T. Logvinenko. A theoretical framework for sequential importance sampling and resampling. In A. Doucet, N. De Freitas, and N. Gordon, editors, *Sequential Monte Carlo Methods in Practice*. Springer, 2001.
- [13] J. S. Liu. Metropolized independent sampling with comparisons to rejection sampling and importance sampling. *Stat. Comput.*, 6:113–119, 1996.
- [14] D. B. Rubin. A noniterative sampling/importance resampling alternative to the data augmentation algorithm for creating a few imputations when the fraction of missing information is modest: the SIR algorithm (discussion of Tanner and Wong). *J. Am. Statist. Assoc.*, 82:543–546, 1987.