Using systematic sampling for approximating Feynman-Kac solutions by Monte Carlo methods

Ivan Gentil* and Bruno Rémillard†

Abstract

While convergence properties of many sampling selection methods can be proven to hold in a context of approximation of Feynman-Kac solutions using sequential Monte Carlo simulations, there is one particular sampling selection method introduced by Baker (1987), closely related with "systematic sampling" in statistics, that has been exclusively treated on an empirical basis. The main motivation of the paper is to start to study formally its convergence properties, since in practice it is by far the fastest selection method available. One will show that convergence results for the systematic sampling selection method are related to properties of peculiar Markov chains.

Keywords: Feynman-Kac formulæ, sequential Monte Carlo, genetic algorithms, systematic sampling, Markov chains.

MSC (2000): 65C35, 60G35, 60J10.

1 Introduction

Let $(X_k)_{k\geq 0}$ be a non-homogeneous Markov chain on a locally compact metric space E, with transition kernels $(K_n)_{n\geq 1}$ and initial law η_0 defined on the Borel σ-field $\mathcal{B}(E)$. Further let $\mathcal{B}_b(E)$ be the set of bounded $\mathcal{B}(E)$-measurable functions.

Given a sequence $(g_n)_{n\geq 1}$ of positive functions in $\mathcal{B}_b(E)$, suppose that one wants to calculate recursively the following Feynman-Kac formulæ $(\eta_n)_{n\geq 1}$:

$$\eta_n(f) = \frac{\gamma_n(f)}{\gamma_n(1)}, \quad f \in \mathcal{B}_b(E), \tag{1}$$

where

$$\gamma_n(f) = E\left(f(X_n) \prod_{k=1}^{n-1} g_k(X_{k-1}) \right). \tag{2}$$

Note that most nonlinear filtering problems are particular cases of Feynman-Kac formulæ.

Following Crisan et al. (1999) and Del Moral and Miclo (2000), let $M_1(E)$ denotes the set of probability measures on $(E, \mathcal{B}(E))$. If $\mu \in M_1(E)$ and $n \geq 0$, let μK_n be the probability measure defined on $\mathcal{B}_b(E)$ by

$$\mu K_n(f) = \mu(K_n f) = \int_E \int_E f(z) K_n(x, dz) \mu(dx).$$

In order to understand the relation between the η_ns, for any $n \geq 1$, let $\psi_n : M_1(E) \mapsto M_1(E)$ be defined by

$$\psi_n(\eta) f = \frac{\eta(g_n f)}{\eta(g_n)}, \quad \eta \in M(E), f \in \mathcal{B}_b(E),$$

and let Φ_n denotes the mapping from $M_1(E)$ to $M_1(E)$ defined by

$$\Phi_n(\eta) = \psi_n(\eta) K_n.$$

*CEREMADE, Université Paris-Dauphine, Place du Maréchal de Lattre de Tassigny, F-75775 Paris cedex 16, France, gentil@ceremade.dauphine.fr

†HEC Montréal, Service de l'enseignement des méthodes quantitatives de gestion, 3000, chemin de la côte-Sainte-Catherine, Canada H3T 2A7, Bruno.Remillard@hec.ca

Then it is easy to check that for any $n \geq 1$,

$$\eta_n = \Phi_n(\eta_{n-1}). \tag{3}$$

Note also that for any $n \geq 1$, the mapping Φ_n can be decomposed into

$$\begin{array}{rl} \hat{\eta}_n = & \psi_{n+1}(\eta_n), \\ \eta_{n+1} = & \hat{\eta}_n K_{n+1}, \end{array} \qquad n \geq 0,\ \eta_0 \in M_1(E). \tag{4}$$

Further remark that the first transformation, $\eta_n \mapsto \hat{\eta}_n$, is non-linear, while the second one, $\hat{\eta}_n \mapsto \eta_{n+1}$, is linear.

Even if the forward system of equations (3) looks simple, it can rarely be solved analytically, and even if this is the case, it would require extensive calculations. This is why algorithms for approximating $(\eta_n)_{n\in\mathbb{N}}$, starting from η_0, are so important.

One such method, presented in the remarkable surveys Del Moral and Miclo (2000), Crisan and Doucet (2002) and the book of Del Moral (2004), is to build approximations of measures $(\eta_n)_{n\in\mathbb{N}}$ using interacting particle systems. The algorithm uses decomposition (4), and by analogy with genetics, the first step, which is related to a sampling selection method, is often referred to as the *selection* step, and the second one is termed the *mutation* step, while in reality it is a Markovian evolution of the particles. The speed of the latter cannot be improved in general, so the speed of any algorithm depends on the rapidity of the sampling selection process.

In this paper, one discusses properties of a particular algorithm that is called "systematic sampling" selection herein, while in the genetic algorithms literature, it has been strangely called "Stochastic universal sampling" selection. It seems to have appeared first in Baker (1987). It has been reintroduced in the filtering literature in Künsch (2005).

In what follows, a description of the general algorithm is given in Section 2, with a few examples of sampling selection methods, together with some tools for studying its convergence. In Section 3, one focuses on the systematic sampling selection method, giving some properties, and stating some convergence results and a conjecture, based on results from Markov chains proved in the appendix. Finally, in Section 4, numerical comparisons between sampling selection methods are made through a simple model of nonlinear filtering for noisy black-and-white images.

2 Algorithm and sampling selection methods

The general algorithm for approximating the solution of (3) is first given, following the exposition in Crisan et al. (1999), Del Moral and Miclo (2000), while particular sampling selection methods are presented next. Throughout the rest of the paper, it is assumed that for any $n \geq 1$, $\inf_{x\in E} g_n(x) > 0$.

2.1 General algorithm

Let N be a integer, representing the number of particles and for any $n \geq 0$, let $\xi_n = \left\{\xi_n^1, \cdots, \xi_n^N\right\}$ denotes the particles at time n and set

$$\eta_n^N = \frac{1}{N}\sum_{i=1}^N \delta_{\xi_n^i}.$$

- At time $n = 0$, the initial particle system $\xi_0 = \left\{\xi_0^1, \cdots, \xi_0^N\right\}$ consist of N independent and identically distributed particles with common law η_0.
- For each $n \geq 1$, the particle system $\xi_n = \left\{\xi_n^1, \cdots, \xi_n^N\right\}$ consists of N particles, is obtained in the following way:

(Sampling/Selection) First calculate the weights vector $W_n \in (0,1)^N$, where

$$W_n^i = \frac{g_n(\xi_{n-1}^i)}{\sum_{i=1}^N g_n(\xi_{n-1}^i)}, \qquad 1 = 1, \ldots, N. \tag{5}$$

Then, select, according to a given sampling selection method, a sample $\hat{\xi}_{n-1} = \left\{\hat{\xi}_{n-1}^1, \ldots, \hat{\xi}_{n-1}^N\right\}$ of size N from ξ_{n-1}.

(Evolution/Mutation) Given $\hat{\xi}_{n-1}$, the new particle system ξ_n consists of particles ξ_n^i chosen independently from law $K_n(\hat{\xi}_{n-1}^i, dx)$, $1 \le i \le N$. In other words, for any $z = (z^1, \ldots, z^N) \in E^N$,

$$P\Big(\xi_n \in dx | \hat{\xi}_{n-1} = z\Big) = \bigotimes_{i=1}^N K_n(z^i, dx^i).$$

Note that in order to describe a sampling selection method, it suffices to define how the numbers $M_n^1, \ldots, M_n^N \in \{0, 1, \ldots, N\}$ are randomly selected, with M_n^i representing the number of times particle ξ_{n-1}^i is appears in the new sample. Therefore, one can write

$$\hat{\eta}_{n-1}^N = \frac{1}{N}\sum_{i=1}^N \delta_{\hat{\xi}_{n-1}^i} = \frac{1}{N}\sum_{i=1}^N M_n^i \delta_{\xi_{n-1}^i}.$$

A sampling selection method will be said to be conditionally unbiased, if for any $i \in \{1, \ldots, N\}$ and any $k \ge 1$, $E(M_k^i | \xi_{k-1}) = N W_k^i$.

Remark 2.1. *Conditional unbiasedness yields the following property:*

$$E\big(\eta_k^N f | \xi_{k-1}\big) = \Phi_k\big(\eta_{k-1}^N\big)(f), \qquad f \in \mathcal{B}_b(E). \tag{6}$$

For, in that case,

$$\begin{aligned} E\big(\eta_n^N f | \xi_{n-1}\big) &= E\Big(E\Big(\eta_n^N f | \xi_{n-1}, \hat{\xi}_{n-1}\Big) | \xi_{n-1}\Big) = E\left(\frac{1}{N}\sum_{i=1}^N M_n^i K_n f(\xi_{n-1}^i) \middle| \xi_{n-1}\right) \\ &= \sum_{i=1}^N W_n^i K_n f(\xi_{n-1}^i) = \Phi_n\big(\eta_{n-1}^N\big)(f). \end{aligned}$$

The mean square error of a particular sampling selection method can be obtained using the following useful result, obtained by (Del Moral and Miclo, 2000, Theorem 2.36). Before stating the result, define, for any measurable η with values in $M(E)$,

$$\|\eta\|_2^2 = \sup_{f \in \mathcal{B}_b(E),\ \|f\|_\infty \le 1} E\left((\eta f)^2\right).$$

Theorem 2.2. *Assume that the sampling selection method is conditionally unbiased and that the following condition is verified for all $1 \le k \le n$: there exists a constant C_k such that for all N-dimensional vectors $\{q^1, \cdots, q^N\} \in \mathbb{R}^N$,*

$$E\left[\left(\frac{1}{N}\sum_{i=1}^N (M_k^i - N W_k^i) q^i\right)^2 \middle| \xi_{k-1}\right] \le \frac{1}{N} C_k \max_{1 \le i \le N} \left|q^i\right|^2. \tag{7}$$

Then, for all $1 \le k \le n$, there exists a constant C_k' such that

$$\|\eta_k^N - \eta_k\|_2^2 \le C_k'/N.$$

In what follows, only conditionally unbiased sampling selection methods are considered. As shown in Remark 3.3, one can see that in general, the systematic sampling selection method defined below does not satisfy condition (7) of the previous theorem, while classical sampling selection methods, like the ones listed in Section 2.3, do satisfy it. Therefore, weaker conditions must be imposed in order to obtain mean square convergence. In fact, one has the following result.

Theorem 2.3. *Let (a_N) be a sequence such that $a_N/N \to 0$, as $N \to \infty$. Assume that the sampling selection method is conditionally unbiased. Then*

$$\lim_{N\to\infty} a_N \max_{1\le k\le n} \|\eta_k^N - \eta_k\|_2^2 = 0$$

if and only if, for any $f \in \mathcal{B}_b(E)$,

$$\lim_{N\to\infty} a_N \max_{1\le k\le n} E\left[\left(\frac{1}{N}\sum_{i=1}^N (M_k^i - NW_k^i) f(\xi_{k-1}^i)\right)^2\right] = 0. \tag{8}$$

Moreover, $\sup_{N\ge1} a_N \max_{1\le k\le n} \|\eta_k^N - \eta_k\|_2^2$ *is finite if and only if*

$$\sup_{N\ge1} a_N \max_{1\le k\le n} E\left[\left(\frac{1}{N}\sum_{i=1}^N (M_k^i - NW_k^i) f(\xi_{k-1}^i)\right)^2\right] < \infty.$$

Proof. Suppose that $a_N/N \to 0$ and let $f \in \mathcal{B}_b(E)$ be given. First, note that using the unbiasedness condition, together with (6), one has, for any $k \in \{1,\ldots,n\}$,

$$E\left[(\eta_k^N f - \eta_k f)^2\right] = E\left[(\eta_k^N f - \Phi_k(\eta_{k-1}^N) f)^2\right] + E\left[(\Phi_k(\eta_{k-1}^N) f - \eta_k f)^2\right]. \tag{9}$$

Since $g_k \ge c_k > 0$ by hypothesis, for some positive constant c_k, $k \ge 1$, it follows that

$$\lim_{N\to\infty} a_N \max_{1\le k\le n} \|\eta_k^N - \eta_k\|_2^2 = 0$$

if and only if for any $k = 1,\ldots,n$, $\lim_{N\to\infty} a_N E\left[(\eta_k^N f - \Phi_k(\eta_{k-1}^N) f)^2\right] = 0$. Next, it can be shown easily that for any $k = 1,\ldots,n$, $E\left[\left(\eta_k^N f - \Phi_n(\eta_{k-1}) f\right)^2 \middle| \xi_{k-1}\right]$ can be written as

$$E\left[\left(\frac{1}{N}\sum_{i=1}^N (M_k^i - NW_k^i) K_k f(\xi_{k-1}^i)\right)^2 \middle| \xi_{k-1}\right] + \frac{1}{N}\Phi_k(\eta_{k-1}^N)(K_k f^2 - (K_k f)^2).$$

Since $K_k f \in \mathcal{B}_b(E)$, $0 \le \frac{1}{N}\Phi_k(\eta_{k-1}^N)(K_k f^2 - (K_k f)^2) \le \frac{1}{N}\|f\|_\infty^2$, and $a_N/N \to 0$, it follows from the calculations above that

$$\lim_{N\to\infty} a_N \max_{1\le k\le n} E\left[(\eta_k^N f - \Phi_k(\eta_{k-1}^N) f)^2\right] = 0$$

if and only if (8) holds true. The rest of the proof is similar, so it is omitted. □

2.2 Systematic sampling

By obvious analogy with systematic sampling in Statistics, the first sampling selection method that is described is simply called "systematic sampling". It appears that this method was first proposed by Baker (1987) under the strange name "Stochastic Universal Sampling", in a context of unbiased sampling selection for genetic algorithms. However, nobody formally studied its convergence properties.

As opposed to the definition of Baker (1987), the sampling selection method can simply be defined in the following way: For $n \geq 1$, let U_n a uniform random variable on $[0,1)$ and note for $w \in [0,1]$, $M(w, U_n) := \lfloor Nw + U_n \rfloor$, where $\lfloor x \rfloor$ denotes the integer part of x. Then

$$\begin{aligned} M_n^1 &:= M(W_n^1, U_n), \\ M_n^k &:= M(W_n^1 + \cdots + W_n^k, U_n) - M(W_n^1 + \cdots + W_n^{k-1}, U_n), \quad k = 2, \ldots, N. \end{aligned}$$

Since $M(1, U_n) = N$, one gets that $\sum_{i=1}^N M_n^i = N$. Therefore the number of particles is always N. Properties of that sampling selection method are examined in Section 3.

2.3 Other sampling methods

One can grossly classify the various sampling selection methods into two categories, according as the number of particles is constant or random. The following list is by no means exhaustive. For the first two methods, N is constant, while N_n fluctuates in the last two methods. For other sampling selection methods, one may consult Crisan et al. (1999), Del Moral and Miclo (2000), Del Moral (2004) and references therein. Note that the last two methods are particular cases of what is known as "Branching selection methods" in the filtering literature.

2.3.1 Simple random sampling

This selection method is based on simple random sampling without rejection. It follows that

$$\left(M_n^1, \cdots, M_n^N\right) \sim \text{Multinomial}\left(N, W_n^1, \cdots, W_n^N\right),$$

where $\left(W_n^i\right)_{1 \leq i \leq N}$ are given by (5). This sampling selection method is computationally demanding, but it has many interesting properties that have been studied mainly by Del Moral and co-authors, e.g. see Del Moral (2004). In particular, conditions (i)–(ii) of Theorem 2.2 are met; also one can prove a Central Limit Theorem and Large Deviations Properties.

2.3.2 The remainder stochastic sampling

This algorithm was first introduced by Brindle (1980) in a context of unbiased sampling selection for genetic algorithms; see also Baker (1985, 1987) for comparisons between sampling selection methods in the latter context. It is also defined as "Residual sampling" by Liu and Chen (1998). It is much faster to implement than the simple random sampling selection method, it satisfies conditions (i)–(ii) of Theorem 2.2, and recently, Douc and Moulines (2005) investigated some of its convergence properties. See also Del Moral and Miclo (2000) and the references therein. To describe the selection method, first define $\tilde{N} = N - \sum_{i=1}^N \lfloor W_n^i \rfloor = \sum_{i=1}^N \left\{NW_n^i\right\}$, where $\{x\}$ stands for the fractional part of x, i.e. $\{x\} = x - \lfloor x \rfloor$. Next, allocate the (possibly) remaining $\tilde{N}$ particles via simple random sampling, i.e.

$$\left(M_n^1 - \lfloor W_n^1 \rfloor, \cdots, M_n^N - \lfloor W_n^N \rfloor\right) \sim \text{Multinomial}\left(\tilde{N}, \tilde{W}_n^1, \cdots, \tilde{W}_n^N\right),$$

with $\tilde{W}_n^i = \left\{NW_n^i\right\} / \sum_{j=1}^N \left\{NW_n^j\right\}$, $1 \leq i \leq N$.

2.3.3 Binomial sampling

As stated before, for this sampling selection method and the next one, the number of particles at time n is random and it is denoted by N_n, $n \geq 0$. Of course, N_0 is fixed. For $n \geq 1$, and given ξ_{n-1} and N_{n-1}, $M_n^1, \ldots, M_n^{N_{n-1}}$ are independent and $M_n^i \sim \text{Bin}(N_{n-1}, W_n^i)$, for $1 = 1, \ldots, N_{n-1}$. It follows that

$$N_n = \sum_{i=1}^{N_{n-1}} M_n^i.$$

This sampling selection method is a little bit faster than the simple random sampling selection method, but a major drawback is that there is no control on the number of particles. Moreover, $P(N_n = 0) > 0$.

2.3.4 Bernoulli sampling

The Bernoulli sampling selection selection method was introduced in Crisan et al. (1998). See also Crisan and Lyons (1999), Crisan (2003) for additional properties of the sampling selection selection. It is worth noting that M_n^i takes the same values as in the systematic sampling selection method, provided $N_{n-1} = N$. In fact, for $n \geq 1$, and given ξ_{n-1} and N_{n-1}, $M_n^1, \ldots, M_n^{N_{n-1}}$ are independent, where M_n^i is defined by

$$M_n^i = \lfloor N_{n-1} W_n^i \rfloor + \varepsilon_n^i, \quad \varepsilon_n^i \sim \mathrm{Ber}(\{N_n W_n^i\}), \quad 1 \leq i \leq N_{n-1}.$$

Note that $N_n \geq 1$ and that the following alternative representation also holds:

$$M_n^i = \lfloor N(W_n^1 + \cdots + W_n^i) + U_n^i \rfloor - \lfloor N(W_n^1 + \cdots + W_n^{i-1}) + U_n^i \rfloor,$$

where $U_n^1, \ldots, U_n^{N_{n-1}}$ are independent and $U_n^i \sim \mathrm{Unif}([0,1))$, given ξ_{n-1}, N_{n-1}.

3 Some properties and results for systematic sampling selection

Throughout the rest of paper, the selection method is the one defined in Section 2.2. Let's start first with some elementary properties of systematic sampling selection.

Lemma 3.1. *Suppose that U_n is uniformly distributed over $[0,1)$. Then, conditionally on ξ_{n-1}, one has, for any $i \in \{1, \cdots, N\}$,*

$$M_n^i - \lfloor N W_n^i \rfloor \sim \mathrm{Ber}(\{N W_n^i\}). \tag{10}$$

In particular, for any $i \in \{1, \cdots, N\}$, $E\big(M_n^i | \xi_{n-1}\big) = N W_n^i$.

Proof. It suffices to show that whenever $U \sim \mathrm{Unif}([0,1))$ and $x, y \geq 0$, then $\lfloor U + x + y \rfloor - \lfloor U + x \rfloor - \lfloor y \rfloor$ is a Bernoulli random variable with parameter $p = \{y\}$. To this end, first note that $V = \{U + x\}$ is also uniformly distributed on $[0,1)$. Next,

$$\lfloor U + x + y \rfloor - \lfloor U + x \rfloor - \lfloor y \rfloor = \lfloor \{U + x\} + \{y\} \rfloor = \lfloor V + \{y\} \rfloor \sim \mathrm{Ber}(\{y\}).$$

Hence the result. ◻

Remark 3.2. *Using the same proof as in Lemma 3.1, then, conditionally on ξ_{n-1}, one obtains $M_n^i + \cdots + M_n^j - \lfloor N(W_n^i + \cdots + W_n^j) \rfloor \sim \mathrm{Ber}\left(\left\{N(W_n^i + \cdots + W_n^j)\right\}\right)$, for any $i \leq j \in \{1, \cdots, N\}$. Note also that since the sampling selection method is unbiased, i.e. condition (i) of Theorem 2.2 is satisfied, then for any $n \geq 1$, one has*

$$E\big(\eta_n^N f | \xi_{n-1}\big) = \Phi_n\big(\eta_{n-1}^N\big)(f), \qquad f \in \mathcal{B}_b(E).$$

To obtain L^2 convergence of the algorithm based on the systematic sampling selection method, one would like to apply Theorem 2.2 of Del Moral and Miclo (2000). All sampling selections presented in Section 2.3 satisfies property (8). If the N_n is random, there is an similiar condition to (8). But as shown next, systematic sampling behaves differently.

Remark 3.3. *Inequality (7) is not verified in general for the systematic sampling selection method. Here is an illustration. Suppose that $N = 2m$ and let, for any $i \in \{1, \cdots, N/2\}$, $W_n^{2i} = 1/(2N)$, and $W_n^{2i-1} = 3/(2N)$. Then one can check that for any $1 \leq i \leq N/2$,*

$$\begin{array}{llll} M_n^{2i-1} & = 1 \text{ if } U_n \in [0,1/2), & M_n^{2i-1} & = 2 \text{ if } U_n \in [1/2,1), \\ M_n^{2i} & = 1 \text{ if } U_n \in [0,1/2), & M_n^{2i} & = 0 \text{ if } U_n \in [1/2,1). \end{array}$$

Next, if $1 \leq i \leq N/2$, set $q^{2i} = 1$ and $q^{2i-1} = -1$. It follows that

$$E\left[\left(\frac{1}{N}\sum_{i=1}^{N}(M_k^i - N W_k^i) q^i\right)^2 \middle| \xi_{k-1}\right] = \frac{1}{4},$$

showing that inequality (7) is false.

However one believes that the following holds true.

Conjecture 3.4. *Suppose that η_0 and $(K_n)_{n\geq 1}$ are absolutely continuous laws and consider that $M_n^1, \ldots, M_n^N$ are obtained using the the systematic sampling selection method. Then, for all $f \in \mathcal{B}_b(E)$ and $n \geq 1$, (8) holds with $a_N \equiv 1$, i.e.*

$$\lim_{N\to\infty} E\left[\left(\frac{1}{N}\sum_{i=1}^{N}(M_n^i - NW_n^i)f(\xi_{n-1}^i)\right)^2\right] = 0. \tag{11}$$

Note that it follows from Theorem 2.3 that the above conjecture is equivalent to $\|\eta_n^N - \eta_n\|_2 \to 0$, as $N \to \infty$, for any $n \geq 0$. In what follows, one tries to motivate why Conjecture 3.4 might be true. To this end, first note that any $1 \leq k \leq N$,

$$M_n^k - NW_n^k = \left\{N(W_n^1 + \cdots + W_n^{k-1}) + U_n\right\} - \left\{N(W_n^1 + \cdots + W_n^k) + U_n\right\}.$$

Now, set $F_n^0 = 0$ and $F_n^k = \sum_{j=1}^{k} g_n(\xi_{n-1}^j)$, $1 \leq k \leq N$. For any $\alpha > 0$ and any $f \in \mathcal{B}_b(E)$, further define

$$\begin{aligned} Z_n^N(f,\alpha) &= \frac{1}{\sqrt{N}}\sum_{k=1}^{N} f(\xi_{n-1}^k)\left(\left\{\frac{F_n^{k-1}}{\alpha} + U_n\right\} - \left\{\frac{F_n^k}{\alpha} + U_n\right\}\right) \\ &= \frac{1}{\sqrt{N}}\sum_{k=1}^{N} f(\xi_{n-1}^k)\left(\{S_n^{k-1}\} - \{S_n^k\}\right), \end{aligned}$$

where $S_n^k = F_n^k/\alpha + U_n$ and $S_n^0 = U_n$. Then, setting $\bar{g}_n = \frac{1}{N}\sum_{k=1}^{N} g_n(\xi_n^k)$ and defining $Y_n^N(f) = Z_n^N(f, \bar{g}_n)$, one has $Y_n^N(f) = \frac{1}{\sqrt{N}}\sum_{i=1}^{N}(M_n^i - NW_n^i)f(\xi_{n-1}^i)$, so one can rewrite (11) in the form

$$\lim_{N\to\infty} \frac{1}{N} E\left[\left(Y_n^N(f)\right)^2\right] = 0.$$

Unfortunately, working with Y_n^N appears to be impossibly difficult, so one could work instead with a more tractable quantity, namely Z_n^N. In the case $n = 1$, one has a least the following result, which is a first step in proving Conjecture 3.4. Before stating it, recall that $D([0,1])$ is the space of càdlàg functions with Skorohod's topology.

Theorem 3.5. *Assume that the law of $\left\{g_1(\xi_0^1)\right\}$ is absolutely continuous. Then, for any $\alpha > 0$ and any $f \in \mathcal{B}_b(E)$, the sequence of processes $B_N \in D([0,1])$ defined by*

$$B_{f,\alpha}^N(t) = Z_1^{\lfloor Nt \rfloor}(f,\alpha), \qquad t \in [0,1],$$

converges in $D([0,1])$ to $\sigma B_{f,\alpha}$, where $B_{f,\alpha}$ is a Brownian motion and

$$\lim_{N\to\infty} E\left[\left(Z_1^N(f,\alpha)\right)^2\right] = \sigma^2.$$

The proof on Theorem 3.5 is an easy consequence of Theorem A.3 applied with $X_k = f(\xi_0^k)$, $Y_k = \left\{g_1(\xi_0^k)\right\}$ and $f(x,y,s) = x(y-s)$. In addition, there is an "explicit" expression for σ^2. More details can be found in Appendix A.

Remark 3.6. *Theorem 3.5 does not prove Conjecture 3.4 in the case $n = 1$. However, if one is willing to deal with a random number of particles at step $n = 1$, one obtains the following interesting result: Set $N_1 = \lfloor U_1 + \sum_{j=1}^{N} g_1(\xi_0^k)/\alpha \rfloor$, and define*

$$\hat{\eta}_0^{N_1} = \frac{1}{N_1} \sum_{k=1}^{N} M_1^k \delta_{\xi_0^k}.$$

Then, as $N \to \infty$, N_1/N tends to $\eta_0(g_1)/\alpha$ and $\limsup_{N\to\infty} N \|\eta_1^{N_1} - \eta_1\|_2^2 < \infty$.

To prove convergence for higher orders, i.e $n > 1$, one would need results from non-homogeneous Markov chains. The approach will be examined in a near future, using for example the results of Sethuraman and Varadhan (2005).

Remark 3.7. *In order to keep N_1 fixed, one could try to control the term $Z_1^N(\overline{g}_1) - Z_1^N(\alpha)$. Since $\sqrt{N}(\overline{g}_1 - \alpha) \rightsquigarrow \sqrt{\eta_0(g_1^2)}\,\mathcal{Z}$, where $\mathcal{Z} \sim \mathcal{N}(0,1)$, it follows that*

$$\sqrt{N}\left[\frac{1}{\overline{g}_1} - \frac{1}{\alpha}\right] = -\frac{\sqrt{\eta_0(g_1^2)}}{\eta_0^2(g_1)}\,\mathcal{Z} + o_P(1).$$

If one could differentiate term by term, one would then obtain

$$\left(Z_1^N(\overline{g}_1) - Z_1^N(\alpha)\right) \rightsquigarrow \eta_0(K_1 f g_1)\frac{\sqrt{\eta_0(g_1^2)}}{\eta_0(g_1)^2}\,\mathcal{Z} = \eta_1(f)\,\frac{\sqrt{\eta_0(g_1^2)}}{\eta_0(g_1)}\,\mathcal{Z},$$

so one could guess that $Y_N \rightsquigarrow \eta_1(f)\frac{\sqrt{\eta_0(g_1^2)}}{\eta_0(g_1)}\,\mathcal{Z} + B_{f,\alpha}(1)$. On the other hand, if the sequence $Z_1^N(\alpha)$ was tight for α in a closed interval not containing zero, then one would get $Z_1^N(\overline{g}_1) - Z_1^N(\alpha) \to 0$ in probability. There is no indication so far in favor of one of these two approaches.

4 Numerical comparisons

The numerical comparions will be done through a simple model of filtering for tracking a moving target using noisy black-and-white images, where the exact filter can be calculated explicitly, that is η_n is known for any $n \geq 1$, e.g. Gentil et al. (2005).

4.1 Description of the model

One will assume that the target moves on $\mathbb{Z}^2$ according to a Markov chain. Observations consist in black-and-white noisy images of a finite fixed region $R \subset \mathbb{Z}^2$. More precisely, let $(X_n)_{n\geq 0}$ be a homogeneous Markovian chain with values in $\mathfrak{X} = \{\omega \in \{0,1\}^{\mathbb{Z}} : \sum_{x\in\mathbb{Z}} \omega(x) = 1\}$. Of course, the position of the target at step n is x_0 if and only if $X_n(x_0) = 1$. Set

$$M(a,b) = P\left\{X_{n+1}(a) = 1 | X_n(b) = 1\right\}, \qquad a, b \in \mathbb{Z}^2. \tag{12}$$

Note that M describes exactly the movement of the target.

The model for observations $Y_k \in \{0,1\}^R$, $k = 1, \ldots, n$ is the following: Given $X_0, \ldots, X_n$, assume that $\{Y_n(x)\}_{x\in R}$ are independent and for any $x \in R$,

$$P(Y_n(x) = 0 | X_n(x) = 0) = p_0, \qquad P(Y_n(x) = 1 | X_n(x) = 1) = p_1, \tag{13}$$

where $0 < p_0, p_1 < 1$. One wants to compute the distribution of X_k conditionally to $\mathcal{Y}_n$, where $\mathcal{Y}_n$ is the sigma-algebra generated by observations $Y_1, \ldots, Y_n$, and $\mathcal{Y}_0$ is the trivial sigma-algebra. As in

Section 2 of Gentil et al. (2005), note that for any $(\omega, \omega') \in \{0,1\}^R \otimes \mathfrak{X}$, the conditional probability $P(Y_k = \omega | X_k = \omega') = \Lambda(\omega, \omega')$ satisfies

$$\Lambda(\omega, \omega') = p_0^{|R|-1}(1-p_1)\left(\frac{1-p_0}{p_0}\right)^{<\omega>}\left(\frac{p_0 p_1}{(1-p_0)(1-p_1)}\right)^{<\omega\omega'>},$$

where $<\omega> = \sum_{x \in R} \omega(x)$ and $<\omega\omega'> = \sum_{x\in R} \omega(x)\omega'(x)$.

Let P be the joint law of the Markovian targets with initial distribution ν, and the observations, and let Q be the joint law of the Markovian targets with initial distribution ν, and independent Bernoulli observations with mean 1/2. Further let $\mathcal{G}_n$ be the sigma-algebra generated by $Y_1, \ldots, Y_n, X_0, \ldots, X_n$. Then it is easy to check that with respect to $\mathcal{G}_n$, P is equivalent to Q and $\left.\frac{dP}{dQ}\right|_{\mathcal{G}_n} = \prod_{j=1}^n 2^{|R|}\Lambda(Y_j, X_j)$. Further define $L_n = \prod_{j=1}^n \Lambda(Y_j, X_j)$. Denoting by E_P (resp. E_Q) expectation with respect to P (resp. Q), observe that for any $f \in \mathcal{B}_b(\mathfrak{X})$, one has

$$\hat{\eta}_n(f) = E_P\left(f(X_n)|\mathcal{Y}_n\right) = \frac{E_Q\left(f(X_n)L_n|\mathcal{Y}_n\right)}{E_Q\left(L_n|\mathcal{Y}_n\right)}. \tag{14}$$

This formula is a consequence of the properties of conditional expectations, and in the context of filtering, (14) is known as the Kallianpur-Stribel formula, e.g. Kallianpur and Striebel (1968).

Denote by K the Markov kernel associated with the Markov chain $(X_n)_{n\geq 0}$ defined by M, as in (12). One can check that η_n and $\hat{\eta}_n$ satisfy (4) with $g_n(x) = \Lambda(y_n, x)$ and $K_n = K$. Note also that in that case, g_n takes only two values which can be assumed to belong to $\mathbb{Q}$ because of rounding errors. It follows from Remark A.4 that

$$\sup_{N_0 \geq 1} E\left[N_1 \|\eta_1^{N_1} - \eta_1\|^2\right] < \infty.$$

The results proved in Section 2 of Gentil et al. (2005) provide an algorithm for computing recursively the exact filter, i.e. the law of X_n given $\mathcal{Y}_n$. In the next section, one will compare the results from the exact filter with those obtained by the Monte Carlo algorithm described in Section 2 with various sampling methods.

4.2 Simulation results

In what follows, R is chosen to be the window of size 100×100 defined by $R = \{0, \ldots 99\}^2$. To makes things simple, the target starts at $(50, 50)$ and it moves according to a simple symmetric random walk, i.e its goes up, down, right or left to the nearest neighbor with probability 1/4. The estimation of the position of the target is taken to be the mean of the various measures. The simulations were performed with $p_0 = 0.9$ and $p_1 = 0.9$, that there are 10% of errors in pixels.

In order to compare the efficiency of the optimal filter (OF) and the samplings methods described in section 2, i.e. simple random sampling (SRS), remainder stochastic sampling (RSS), systematic sampling (SyS), binomial sampling (BiS), and Bernoulli sampling (BeS), the mean absolute error between the estimated position and the true one was calculated over several time intervals, namely $[2, 100]$, $[10, 100]$ and $[30, 100]$.

The number of particles N_0 takes values 1000, 10000, 30000 and 50000. The results are reported in the Table 1.

According to these results, one may conclude that the algorithm based on the systematic sampling selection method performs quite well, provided the number of particles is large enough. Surprisingly, the Monte Carlo based approximate filters seem to perform better than the optimal filter. However the difference may not be statistically significant. Next, based on the results of Table 1 for the time interval $[30, 100]$, note that when the target is precisely detected, the error seems to stabilize near zero, indicating that of η_n^N to η_n might be uniform on n. Finally, other simulations performed with several moving targets seem to indicate that the algorithm based on systematic sampling also give impressive results.

Table 1: Mean absolute error for one target performing a simple symmetric random walk in images of size 100×100 with 10% of errors.

	t	$[2, 100]$	$[10, 100]$	$[30, 100]$
	OF	2.1	1.7	1.4
$N_0 = 1000$	SRS	57.4	60.3	56.2
	RSS	51.8	53.6	43.8
	SyS	42.7	43.7	36.9
	BiS	54.0	56.5	45.5
	BeS	13.8	12.1	6.9
$N_0 = 10000$	SRS	76.0	81.0	64.1
	RSS	1.9	0.8	0.5
	SyS	2.4	0.7	0.5
	BiS	6.4	6.7	0.5
	BeS	77.5	82.5	85.4
$N_0 = 30000$	SRS	8.7	3.0	0.5
	RSS	2.4	0.6	0.4
	SyS	3.9	1.5	0.4
	BiS	4.0	2.1	0.5
	BeS	8.1	6.2	0.4
$N_0 = 50000$	SRS	3.9	3.4	0.9
	RSS	10.2	5.2	0.7
	SyS	5.0	2.5	0.6
	BiS	4.8	2.1	0.8
	BeS	3.6	1.5	0.3

A Convergence results for a Markov chain

Suppose that $(X_i, Y_i)_{i\geq 1}$ are independent observations of $(X,Y) \in \mathfrak{Z} := \mathbb{R}\times[0,1)$ of law P, with marginal distributions P_X and P_Y respectively. Further let λ denotes Lebesgue's measure on $[0,1)$. Given $Z_0 = (X_0, S_0) \in \mathbb{R}\times[0,1)$, set $Z_i = (X_i, \{S_i\})$, where $S_i = S_{i-1} + Y_i$, $i \geq 1$.

For $n \in \mathbb{Z}$, set $e_n(s) = e^{2\pi i n s}$, $s \in [0,2)$, and let $\zeta_n = E\left(e_n(Y)\right)$. Further set $\mathcal{N} = \{n \in \mathbb{Z}; \zeta_n = 1\}$. Recall that $(e_n)_{\in\mathbb{Z}}$ is a complete orthonormal basis of the Hilbert space $H = L^2([0,1),\lambda)$ with scalar product $(f,g) = \int_0^1 f(s)\bar{g}(s)ds$ and norm $\|f\|_2 = \sqrt{(f,f)}$.

It is easy to check that $(Z_i)_{i\geq 0}$ is a Markov chain on $\mathfrak{Z}$ with kernel $\mathcal{K}$ defined by

$$\mathcal{K}f(x,s) := \int_{\mathfrak{Z}} f(x', \{s+y\})P(dx', dy), \quad f \in \mathcal{B}_b(\mathfrak{Z}), \tag{15}$$

and stationary distribution $\mu = P_X \otimes \lambda$. Note that for any $f \in L^2(\mu)$, by Tonelli's theorem, $\mathcal{K}f$ is well defined, it depends only on $s \in [0,1)$, and it belongs to H since

$$\begin{aligned}
\|\mathcal{K}f\|_2^2 &\leq \int_0^1 \int_{\mathfrak{Z}} f^2(x,\{s+y\})P(dx,dy)ds = \int_{\mathfrak{Z}}\int_0^1 f^2(x,u)duP(dx,dy) \\
&= \int_{\mathfrak{Z}} f^2(z)\mu(dz) = \|f\|^2_{L^2(\mu)}.
\end{aligned}$$

Finally, let $\mathcal{L}$ and $\mathcal{A}$ be the linear bounded operators from $L^2(\mu)$ to H defined by

$$\mathcal{L}f(s) = \sum_{n\in\mathcal{N}} (Kf, e_n)e_n(s), \qquad \mathcal{A}f(s) = \int_{\mathbb{R}} f(x,s)P_X(dx), \qquad s \in [0,1).$$

Theorem A.1. *Let $f \in L^2(\mu)$ be given and set $W_N = \dfrac{1}{N}\sum_{k=1}^N f(Z_k)$. Then:*

(i) If the initial distribution of $Z_0 = (X_0, S_0)$ is μ, then W_N converges almost surely and in mean square to $\mathcal{W}$ given by

$$\mathcal{W} = \mathcal{L}f(S_0) = \sum_{n\in\mathcal{N}} (\mathcal{K}f, e_n)e_n(S_0). \tag{16}$$

If, in addition,

$$\sum_{n\in\mathbb{Z}\setminus\mathcal{N}} \frac{|(\mathcal{K}f,e_n)|\,|(\mathcal{A}f,e_n)|}{|1-\zeta_n|} < \infty, \tag{17}$$

then $NE\left[(W_N - \mathcal{W})^2\right]$ converges, as $N\to\infty$, to

$$\|f\|^2_{L^2(\mu)} - \|\mathcal{L}f\|_2^2 + 2\sum_{n\in\mathbb{Z}\setminus\mathcal{N}} \frac{(\mathcal{K}f,e_n)\overline{(\mathcal{A}f,e_n)}}{1-\zeta_n}. \tag{18}$$

(ii) If the initial distribution of $Z_0 = (X_0,S_0)$ is μ, if $\mathcal{N} = \{0\}$ and

$$\sum_{n\in\mathbb{Z}\setminus\{0\}} \frac{|(\mathcal{K}f,e_n)|^2}{|1-\zeta_n|^2} < \infty, \tag{19}$$

then the sequence of processes B_N, defined by $B_N(t) = \sqrt{N}\left(W_{\lfloor Nt\rfloor} - \mu(f)\right)$, $t\in[0,1]$, converges in $D([0,1])$ to σB, where B is a Brownian motion and σ^2 is given by (18).

(iii) If P_Y admits a square integrable density h, then the Markov chain is geometrically ergodic, that is, there exists $\rho \in (0,1)$ such that for any $f \in L^2(\mu)$,

$$|\mathcal{K}^n f(Z_0) - \mu(f)| \le \|h\|_2 \rho^{n-2} \|f\|_{L^2(\mu)}, \quad n \ge 2.$$

Proof. For simplicity, set $\psi = \mathcal{K} f \in H$. To prove (i), start the Markov chain from μ and denote the law of the chain by Q. Then the sequence $(Z_n)_{n\ge 0}$ is stationary, and Birkhoff's ergodic theorem, e.g. (Durrett, 1996, Section 6.2), can be invoked to claim that W_N converges almost surely and in mean square to some random variable $\mathcal{W}$. To show that $\mathcal{W}$ is indeed given by (22), it suffices to show that $E\left[(W_N - \mathcal{W})^2\right]$ tends to 0, as $N \to \infty$. First, note that $E(\mathcal{W}^2) = \|\mathcal{L}f\|_2^2 = \sum_{n\in\mathcal{N}} |(\psi, e_n)|^2$. Next, set $\varphi(s) = \mathcal{A}f(s)$. If $n \in \mathcal{N}$, then $e_n(Y) = 1$ P-a.s., and it follows, by Fubini's theorem, that

$$\begin{aligned}
(\psi, e_n) &= \int_0^1 \int_{\mathfrak{Z}} f(x, \{s+y\}) \overline{e_n(s)} P(dx, dy) ds \\
&= \int_0^1 \int_{\mathfrak{Z}} f(x, u) \overline{e_n(u)} P(dx, dy) du = (\varphi, e_n).
\end{aligned}$$

As a result, $E(\mathcal{W}^2) = \sum_{n\in\mathcal{N}} |(\varphi, e_n)|^2$. Next, using the fact that for any $k \in \mathbb{Z}$, and any $s, y \in [0,1)$, one has $e_k(\{s+y\}) = e_k(s+y) = e_k(s)e_k(y)$, and it follows that

$$\mathcal{K} e_k(s) = \int_{\mathfrak{Z}} e_k(\{s+y\}) P(dx, dy) = \int_{[0,1)} e_k(\{s+y\}) P_Y(dy) = \zeta_k e_k(s), \qquad s \in [0,1).$$

Hence, for any $k \ge 1$ and any $n \in \mathbb{Z}$, one obtains

$$\mathcal{K}^k e_n = \zeta_n^k e_n. \tag{20}$$

Now, using the Markov property of the chain together with (20), one has

$$\begin{aligned}
E(W_N \mathcal{W}) &= E(W_N \overline{\mathcal{W}}) = \frac{1}{N} \sum_{k=1}^N \sum_{n\in\mathcal{N}} \overline{(\psi, e_n)} E\left[f(Z_k)\overline{e_n(S_0)}\right] \\
&= \frac{1}{N} \sum_{k=1}^N \sum_{n\in\mathcal{N}} \overline{(\psi, e_n)} E\left[\mathcal{K}^k f(S_0) \overline{e_n(S_0)}\right] = \frac{1}{N} \sum_{k=1}^N \sum_{n\in\mathcal{N}} \overline{(\psi, e_n)} \left(\mathcal{K}^{k-1}\psi, e_n\right) \\
&= \frac{1}{N} \sum_{j\in\mathbb{Z}} \sum_{k=1}^N \sum_{n\in\mathcal{N}} (\psi, e_j) \overline{(\psi, e_n)} \left(\mathcal{K}^{k-1} e_j, e_n\right) \\
&= \frac{1}{N} \sum_{j\in\mathbb{Z}} \sum_{k=1}^N \sum_{n\in\mathcal{N}} (\psi, e_j) \overline{(\psi, e_n)} \zeta_j^{k-1} (e_j, e_n) \\
&= \frac{1}{N} \sum_{k=1}^N \sum_{n\in\mathcal{N}} |(\psi, e_n)|^2 \zeta_n^{k-1} = E(\mathcal{W}^2),
\end{aligned}$$

since, by definition, $\zeta_n = 1$, for any $n \in \mathcal{N}$.

Next, using stationarity, the Markov property, (20), and also using identity

$$\frac{2}{N^2} \sum_{k=1}^{N-1} \sum_{j=1}^{k} z^{j-1} = \frac{N-1}{1-z} - \frac{z - z^N}{(1-z)^2}, \qquad z \in \mathbb{C}, z \ne 1,$$

it follows that

$$\begin{aligned}
E(W_N^2) &= \frac{1}{N}E\left[f^2(Z_0)\right] + \frac{2}{N^2}\sum_{k=1}^{N-1}\sum_{j=1}^{k} E\left[\mathcal{K}^{j-1}\psi(S_0)f(Z_0)\right] \\
&= \frac{1}{N}\|f\|^2_{L^2(\mu)} + \frac{2}{N^2}\sum_{k=1}^{N-1}\sum_{j=1}^{k}\left(\mathcal{K}^{j-1}\psi,\varphi\right) \\
&= \frac{1}{N}\|f\|^2_{L^2(\mu)} + \sum_{n\in\mathbb{Z}}(\psi,e_n)\overline{(\varphi,e_n)}\left[\frac{2}{N^2}\sum_{k=1}^{N-1}\sum_{j=1}^{k}\zeta_n^{j-1}\right] \\
&= \frac{1}{N}\|f\|^2_{L^2(\mu)} + \frac{N-1}{N}\sum_{n\in\mathcal{N}}(\psi,e_n)\overline{(\varphi,e_n)} \\
&\quad + \frac{2}{N^2}\sum_{n\in\mathbb{Z}\setminus\mathcal{N}}(\psi,e_n)\overline{(\varphi,e_n)}\left[\frac{N-1}{1-\zeta_n} - \frac{\zeta_n-\zeta_n^N}{(1-\zeta_n)^2}\right] \\
&= \frac{1}{N}\|f\|^2_{L^2(\mu)} + \frac{N-1}{N}E(\mathcal{W}^2) \\
&\quad + \frac{2}{N^2}\sum_{n\in\mathbb{Z}\setminus\mathcal{N}}(\psi,e_n)\overline{(\varphi,e_n)}\left[\frac{N-1}{1-\zeta_n} - \frac{\zeta_n-\zeta_n^N}{(1-\zeta_n)^2}\right].
\end{aligned}$$

Collecting the expressions obtained for $E(W_N^2)$ and $E(W_N\mathcal{W})$, one gets

$$\begin{aligned}
E\left[(W_N-\mathcal{W})^2\right] &= \frac{1}{N}\|f\|^2_{L^2(\mu)} - \frac{1}{N}E(\mathcal{W}^2) \\
&\quad + \frac{2}{N^2}\sum_{n\in\mathbb{Z}\setminus\mathcal{N}}(\psi,e_n)\overline{(\varphi,e_n)}\left[\frac{N-1}{1-\zeta_n} - \frac{\zeta_n-\zeta_n^N}{(1-\zeta_n)^2}\right].
\end{aligned} \tag{21}$$

Since $\sum_{n\in\mathbb{Z}\setminus\mathcal{N}}|(\psi,e_n)||(\varphi,e_n)|$ is finite,

$$\sup_{n\mathbb{Z}\setminus\mathcal{N}}\left|\frac{N-1}{1-\zeta_n} - \frac{\zeta_n-\zeta_n^N}{(1-\zeta_n)^2}\right| = \left|\sum_{k=1}^{N-1}\sum_{j=1}^{k}\zeta_n^{j-1}\right| \le \frac{N^2}{2},$$

it follows from (21) and the Dominated Convergence Theorem that

$$\lim_{N\to\infty} E\left[(W_N-\mathcal{W})^2\right] = 0$$

and under the additional condition (17), one also obtains

$$\lim_{N\to\infty} NE\left[(W_N-\mathcal{W})^2\right] = \|f\|^2_{L^2(\mu)} - E(\mathcal{W}^2) + 2\sum_{n\in\mathbb{Z}\setminus\mathcal{N}}\frac{(\psi,e_n)\overline{(\varphi,e_n)}}{1-\zeta_n},$$

completing the proof of (i).

The proof of (ii) is inspired by Durrett (1996). First, note that since $\mathcal{N}=\{0\}$, $\mathcal{L}f=\mu(f)$ for any $f\in L^2(\mu)$ and it follows from (i) that $\frac{1}{N}\sum_{k=1}^{N} f(Z_k)$ converges almost surely and in L^p to $\mu(f)$, for any $1\le p\le 2$. Moreover given any $f\in L^1(\mu)$, one can find $f_n\in L^2(\mu)$ such that $\|f-f_n\|_{L^1(\mu)} < \frac{1}{n}$. It follows that for any $n\ge 1$,

$$\limsup_{N\to\infty} E\left[\left|\frac{1}{N}\sum_{k=1}^{N} f(Z_k)-\mu(f)\right|\right] \le \frac{2}{n} + \limsup_{N\to\infty} E\left[\left|\frac{1}{N}\sum_{k=1}^{N} f_n(Z_k)-\mu(f_n)\right|\right] = \frac{2}{n}.$$

Since the latter is true for any $n \geq 1$, one may conclude that $\frac{1}{N}\sum_{k=1}^{N} f(Z_k)$ converges in L^1 to $\mu(f)$. By Birkhoff's ergodic theorem, $\frac{1}{N}\sum_{k=1}^{N} f(Z_k)$ converges almost surely to $\mu(f)$. Next, let D be the subset of H defined by

$$D = \left\{ h \in H; \sum_{n\in\mathbb{Z}\setminus\{0\}} \frac{|(h,e_n)|^2}{|1-\zeta_n|^2} < \infty \right\},$$

and let Ξ be the operator from D to H that satisfies

$$\Xi h = \sum_{n\in\mathbb{Z}\setminus\{0\}} \frac{(h,e_n)}{1-\zeta_n} e_n.$$

Note that since $(I-\mathcal{K})\Xi h = (I-\mathcal{L})h$, then $\Xi = (I-\mathcal{K})^{-1}(I-\mathcal{L})$ on D. Let $\mathcal{D}$ be the set of all $f \in L^2(\mu)$ such that f satisfies (19), i.e. $\mathcal{K}f \in D$. Then Ξ can be extended to a mapping from $\mathcal{D}$ to $\mathrm{L}^2(\mu)$ viz. $\Xi f = (I-\mathcal{L})f + \Xi\mathcal{K}f$. Using $\mathcal{KL} = \mathcal{LK} = \mathcal{L}$, one obtains that $\Xi = (I-K)^{-1}(I-\mathcal{L})$ on $\mathcal{D}$. Next, if $f \in \mathcal{D}$, set $g = \Xi f$. Since $\mathcal{L}f = \mu(f)$, it follows that

$$\sqrt{N}(W_N - \mu(f)) = \frac{1}{\sqrt{N}}\sum_{k=1}^{N}\left[g(Z_k) - \mathcal{K}g(\{S_{k-1})\}\right] + \frac{1}{\sqrt{N}}\mathcal{K}g(S_0) - \frac{1}{\sqrt{N}}\mathcal{K}g(\{S_N\}).$$

Now, setting $\mathcal{F}_k = \sigma\{Z_j; j \leq k\}$, the terms $\xi_k = g(Z_k) - \mathcal{K}g(\{S_{k-1})\}$ are square integrable martingale differences with respect to $(\mathcal{F}_j)_{j\geq 0}$, i.e. $E(\xi_k|\mathcal{F}_{k-1}) = 0$, and because g^2 and $(\mathcal{K}g)^2$ both belong to $L^1(\mu)$, it follows from (i), as shown above, that

$$\frac{1}{N}\sum_{k=1}^{N} E\left[\xi_k^2|\mathcal{F}_{k-1}\right] = \frac{1}{N}\sum_{k=1}^{N}\left[\mathcal{K}g^2(\{S_{k-1}\}) - (\mathcal{K}g)^2(\{S_{k-1}\})\right]$$

converges almost surely to $\mu(g^2) - \mu((\mathcal{K}g)^2)$. Note that since $\mathcal{K}g = \Xi\mathcal{K}f$, one has $(\mathcal{K}g, \mathcal{L}f) = 0$ and expression (18) can be written as

$$\begin{aligned}
\sigma^2 &= \|(I-\mathcal{L})f\|^2_{L^2(\mu)} + 2(\Xi\mathcal{K}f, \mathcal{A}f) = \|(I-\mathcal{K})g\|^2_{L^2(\mu)} + 2(\mathcal{K}g, \mathcal{A}f)\\
&= \|(I-\mathcal{K})g\|^2_{L^2(\mu)} + 2(\mathcal{K}g, \mathcal{A}(I-\mathcal{L})f) = \|(I-\mathcal{K})g\|^2_{L^2(\mu)} + 2(\mathcal{K}g, \mathcal{A}(I-\mathcal{K})g)\\
&= \mu(g^2 - 2g\mathcal{K}g + (\mathcal{K}g)^2) + 2\mu(g\mathcal{K}g) - 2\mu((\mathcal{K}g)^2) = \mu(g^2) - \mu((\mathcal{K}g)^2).
\end{aligned}$$

Finally, because of the stationarity of $(\xi_k)_{k\geq 1}$, it follows that for any $\epsilon > 0$,

$$\frac{1}{N}\sum_{k=1}^{N} E\left[\xi_k^2 \mathbb{I}(|\xi_k| > \epsilon\sqrt{N})\right] = E\left[\xi_1^2 \mathbb{I}(|\xi_1| > \epsilon\sqrt{N})\right] \to 0,$$

as $N \to \infty$. The conditions of Theorem 7.4 in Durrett (1996) are all met, so one may safely conclude that defining the process $B_N(t) = \sqrt{N}\left(W_{\lfloor Nt\rfloor} - \mu(f)\right)$, $t \in [0,1]$, then B_N converges in $D[0,1]$ to σB, where B is a Brownian motion.

To prove part (iii), note first that since the density h of Y is square integrable, then $\mathcal{N} = \{0\}$, $\sup_{n\geq 1}|\zeta_n| = \rho < 1$, $\zeta_n = (e_n, h)$, and $\|h\|_2^2 = \sum_{n\in\mathbb{Z}}|\zeta_n|^2$. Therefore, for any $g \in H$, $\sum_{n\in\mathbb{Z}}|(g,e_n)|\,|\zeta_n| \leq \|g\|_2\|h\|_2 < \infty$. It follows that for any $k \geq 2$,

$$\mathcal{K}^k f = \mathcal{K}^{k-1}\psi = \sum_{n\in\mathbb{Z}}(\psi, e_n)\zeta_n^{k-1}e_n,$$

the latter series converging absolutely. Thus

$$\begin{aligned}
\sup_{z_0\in\mathfrak{Z}}\left|\mathcal{K}^k f(s) - \mu(f)\right| &= \sup_{s\in[0,1)}\left|\mathcal{K}^{k-1}\psi(s) - \lambda(\psi)\right| \leq \sum_{n\in\mathbb{Z}\setminus\{0\}}|(\psi,e_n)||\zeta_n|\rho^{k-2}\\
&\leq \|h\|_2\|f\|_{L^2(\mu)}\rho^{k-2}.
\end{aligned}$$

This completes the proof of the theorem. $\square$

Remark A.2. *Note that if $\zeta_n = 1$ for some $n > 0$, then $k \mapsto \zeta_k$ is n-periodic, so $\{\zeta_k; n \in \mathbb{Z} \setminus \mathcal{N}\}$ is finite. Therefore $\sup_{k\in\mathbb{Z}\setminus\mathcal{N}} |\zeta_k| = \rho < 1$ and condition (17) is satisfied. Also, if P_Y has a non degenerate absolutely continuous part, then $\mathcal{N} = \{0\}$ and $\sup_{n\geq 1} |\zeta_n| = \rho < 1$, so condition (17) holds true.*

The next result is a straightforward extension of the previous theorem. Before stating it, denote by ν the joint law of (Z_1, S_0), where $S_0 \sim \mathrm{Unif}([0,1))$.

Theorem A.3. *Suppose that $f \in L^2(\nu)$ and set $W_N = \dfrac{1}{N}\displaystyle\sum_{k=1}^{N} f(Z_k, \{S_{k-1}\})$. Then:*

(i) If the initial distribution of $Z_0 = (X_0, S_0)$ is μ, then W_N converges almost surely and in mean square to $\mathcal{W}$ given by

$$\mathcal{W} = \mathcal{L}f(S_0) = \sum_{n\in\mathcal{N}} (\mathcal{K}f, e_n) e_n(S_0), \tag{22}$$

where

$$\mathcal{K}f(s) = \int_{\mathfrak{Z}} f(x, \{s+y\}, s) P(dx, dy).$$

If $\mathcal{A}f(s) = \int_{\mathfrak{Z}} f(x, s, \{s-y\}) P(dx, dy)$, $s \in [0,1)$ and if in addition,

$$\sum_{n\in\mathbb{Z}\setminus\mathcal{N}} \frac{|(\mathcal{K}f, e_n)|\,|(\mathcal{A}f, e_n)|}{|1-\zeta_n|} < \infty, \tag{23}$$

then $NE\left[(W_N - \mathcal{W})^2\right]$ converges, as $N \to \infty$, to

$$\|f\|^2_{L^2(\nu)} - \|\mathcal{L}f\|_2^2 + 2\sum_{n\in\mathbb{Z}\setminus\mathcal{N}} \frac{(\mathcal{K}f, e_n)\overline{(\mathcal{A}f, e_n)}}{1-\zeta_n}. \tag{24}$$

(ii) If $\mathcal{N} = \{0\}$, if the initial distribution of Z_0 is μ and

$$\sum_{n\in\mathbb{Z}\setminus\{0\}} \frac{|(\mathcal{K}f, e_n)|^2}{|1-\zeta_n|^2} < \infty, \tag{25}$$

then the sequence of processes B_N, defined by $B_N(t) = \sqrt{N}\left(W_{\lfloor Nt\rfloor} - \mu(f)\right)$, $t \in [0,1]$, converges in $D([0,1])$ to σB, where B is a Brownian motion and σ^2 is given by (24).

(iii) If P_Y admits a square integrable density h, then the Markov chain is geometrically ergodic, that is, there exists $\rho \in (0,1)$ such that for any $f \in L^2(\mu)$,

$$|\mathcal{K}^n f(Z_1, S_0) - \mu(f)| \leq \|h\|_2 \rho^{n-2} \|f\|_{L^2(\mu)}, \quad n \geq 2.$$

Remark A.4. *For example, suppose that X_k is bounded and set $f(x,y,s) = x(y-s)$. Then it is easy to check that for any $n \in \mathcal{N}$,*

$$\begin{aligned}(\mathcal{K}f, e_n) &= \int_{\mathfrak{Z}\times[0,1)} x(\{y+s\} - s) P(dx, dy)\overline{e_n(s)} ds \\ &= \int_{\mathfrak{Z}\times[0,1)} xu(e_n(y) - 1)\overline{e_n(u)} P(dx, dy) du = 0,\end{aligned}$$

since $P(e_n(Y) = 1) = 1$. It follows from Theorem A.3 that

$$W_N = \frac{1}{N}\sum_{k=1}^{N} X_k(\{S_k\} - \{S_{k-1}\})$$

converges to 0 *almost surely and in mean square.*

Furthermore, if $\mathrm{card}(\mathcal{N}) > 1$ *then condition* (23) *holds and* $\sup_{N\geq 1} NE(W_N^2) < \infty$, *while if* P_Y *is absolutely continuous, then condition* (25) *holds true and* $\sqrt{N}T_N$ *converges in law to a centered Gaussian random variable with variance* σ^2 *given by* (24).

References

Baker, J. E. (1985). Adaptative selection methods for genetic algorithms. In *Proc. International Conf. on Genetic Algorithms amd their Applications.*, pages 101–111. L. Erlbaum Associates.

Baker, J. E. (1987). Reducing bias and inefficiency in the selection algorithm. In *Proc. of the Second International Conf. on Genetic Algorithms amd their Applications.*, pages 14–21. L. Erlbaum Associates.

Brindle, A. (1980). *Genetic algorithms for function optimization.* PhD thesis.

Crisan, D. (2003). Exact rates of convergence for a branching particle approximation to the solution of the Zakai equation. *Ann. Probab.*, 31(2):693–718.

Crisan, D., Del Moral, P., and Lyons, T. (1999). Discrete filtering using branching and interacting particle systems. *Markov Process. Related Fields*, 5(3):293–318.

Crisan, D. and Doucet, A. (2002). A survey of convergence results on particle filtering methods for practitioners. *IEEE Trans. Signal Process.*, 50(3):736–746.

Crisan, D., Gaines, J., and Lyons, T. (1998). Convergence of a branching particle method to the solution of the Zakai equation. *SIAM J. Appl. Math.*, 58(5):1568–1590 (electronic).

Crisan, D. and Lyons, T. (1999). A particle approximation of the solution of the Kushner-Stratonovitch equation. *Probab. Theory Related Fields*, 115(4):549–578.

Del Moral, P. (2004). *Feynman-Kac formulae. Genealogical and interacting particle systems with applications.* Probability and Its Applications. New York, NY: Springer. xviii, 555 p.

Del Moral, P. and Miclo, L. (2000). Branching and interacting particle systems approximations of Feynman-Kac formulae with applications to non-linear filtering. In *Séminaire de Probabilités, XXXIV*, volume 1729 of *Lecture Notes in Math.*, pages 1–145. Springer, Berlin.

Douc, R. and Moulines, E. (2005). Limit theorems for weighted samples with applications to sequential monte carlo methods. preprint.

Durrett, R. (1996). *Probability: theory and examples.* Duxbury Press, Belmont, CA, second edition.

Gentil, I., Rémillard, B., and Del Moral, P. (2005). Filtering of images for detecting multiple targets trajectories. In *Statistical Modeling and Analysis for Complex Data Problem.* Kluwer Academic Publishers, Springer.

Kallianpur, G. and Striebel, C. (1968). Estimation of stochastic systems: Arbitrary system process with additive white noise observation errors. *Ann. Math. Statist.*, 39:785–801.

Künsch, H. R. (2005). Recursive Monte Carlo Filters: Algorithms and Theoretical Analysis. *Ann. Statist.*, 33(5):in press.

Liu, J. S. and Chen, R. (1998). Sequential Monte Carlo methods for dynamic systems. *J. Amer. Statist. Assoc.*, 93(443).

Sethuraman, S. and Varadhan, S. R. S. (2005). A martingale proof of Dobrushin's theorem for non-homogeneous Markov chains. *Electron. J. Probab.*, 10:1221–1235.