Journal of Theoretical Probability, Vol. 13, No. 1, 2000

Convergence of Empirical Processes for Interacting Particle Systems with Applications to Nonlinear Filtering

P. Del Moral[1] and M. Ledoux[1]

Received July 8, 1998

In this paper, we investigate the convergence of empirical processes for a class of interacting particle numerical schemes arising in biology, genetic algorithms and advanced signal processing. The Glivenko–Cantelli and Donsker theorems presented in this work extend the corresponding statements in the classical theory and apply to a class of genetic type particle numerical schemes of the nonlinear filtering equation.

KEY WORDS: Empirical processes; Interacting particle systems; Glivenko–Cantelli and Donsker theorems.

1. INTRODUCTION

1.1. Background and Motivations

Let E be a Polish space endowed with its Borel σ-field $\mathbf{B}(E)$. We denote by $\mathbf{M}_1(E)$ the space of all probability measures on E equipped with the weak topology. We recall that the weak topology is generated by the bounded continuous functions on E and we denote by $\mathscr{C}_b(E)$ the space of these functions. Let $\phi_n: \mathbf{M}_1(E) \to \mathbf{M}_1(E)$, $n \geqslant 1$, be a sequence of continuous functions. Starting from this family we can consider an interacting N-particle system $\xi_n = (\xi_n^1, ..., \xi_n^N)$, $n \geqslant 0$, which is a Markov process with state space E^N and transition probability kernels

$$P(\xi_n \in dx \mid \xi_{n-1} = z) = \prod_{p=1}^{N} \phi_n\left(\frac{1}{N} \sum_{q=1}^{N} \delta_{z^q}\right)(dx^p) \tag{1.1}$$

[1] UMR C55830, CNRS, Université Paul-Sabatier, 31062 Toulouse, France. E-mail: delmoral {ledoux} @cict.fr.

0894-9840/00/0100-0225$18.00/0

where $dx \stackrel{\text{def}}{=} dx^1 \times \cdots \times dx^N$ is an infinitesimal neighborhood of the point $x = (x^1,..., x^N) \in E^N$, $z = (z^1,..., z^N) \in E^N$ and δ_a stands for the Dirac measure at $a \in E$.

Let us introduce the empirical distributions π_n^N of the N-particle system ξ_n

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

which is a random measure on E. Assume that $\xi_0 = (\xi_0^1,..., \xi_0^N)$ is a sequence of N-independent variables with common law $\pi_0 \in \mathbf{M}_1(E)$. Under rather general assumptions, it can be shown that π_n^N converges weakly to a non-random probability distribution $\pi_n \in \mathbf{M}_1(E)$ as $N \to \infty$ [see Del Moral;[1] and Del Moral and Guionnet[15]] and that $\{\pi_n;\, n \geqslant 0\}$ solves the following nonlinear measure valued process equation

$$\pi_n = \phi_n(\pi_{n-1}), \qquad n \geqslant 1$$

This class of interacting particle schemes has been introduced in Del Moral[12, 13] to solve numerically the so-called nonlinear filtering equation. The study of the convergence for abstract functions ϕ_n was initiated in Del Moral,[11] whereas large deviation principles and the associated fluctuations are presented in Del Moral and Guionnet.[14, 15] As described in these papers, under certain conditions on the mappings $\{\phi_n, n \geqslant 1\}$, versions of the law of large numbers and the central limit theorem in the sense of convergence of finite dimensional distributions are available. Namely, for any bounded test function $f: E \to \mathbb{R}$,

$$\pi_n^N(f) \xrightarrow[N \to \infty]{} \pi_n(f) \quad P.\text{ a.s.}$$

and

$$W_n^N(f) = \sqrt{N}(\pi_n^N(f) - \pi(f)) \xrightarrow[N \to \infty]{} W_n(f)$$

where W_n is a centered Gaussian field.

The aim of this paper is to make these two statements uniform over a class of functions $\mathscr{F}$, following the theory of empirical processes for independent samples [cf. van der Vaart and Wellner[38] and the references therein]. Our results apply to a class of measure valued processes and genetic type interacting particle systems arising in biology, in the theory of genetic type interacting particle systems and in advanced signal processing. We prove in this setting Glivenko–Cantelli and Donsker theorems under entropy conditions, as well as exponential bounds for Vapnik–Cervonenkis classes of sets or functions. We provide finally some uniform bounds in both space and time for a class of processes without memory.

1.2. Description of the Model and Statements of the Results

1.3. Description of the Model

To describe precisely our model, we need first introduce some notations. For any Markov transition kernel K and any probability measure μ on E, we denote by μK the probability measure defined by, for any $f \in \mathscr{C}_b(E)$,

$$\mu K(f) = \int \mu(dx)\, K(x, dz)\, f(z)$$

Our measure valued dynamical system is then described by the equation

$$\pi_n = \phi_n(\pi_{n-1}), \qquad n \geqslant 1 \tag{1.2}$$

where

$$\phi_n(\mu) = \psi_n(\mu)\, K_n, \qquad \psi_n(\mu)(f) = \frac{\mu(g_n f)}{\mu(g_n)}$$

and

- $\{K_n; n \geqslant 1\}$ is a sequence of Markov kernels on E,
- $\{g_n; n \geqslant 1\}$ is a collection of bounded positive functions on E such that, for any $n \geqslant 1$, there exist $a_n \in [1, \infty)$ with

$$\forall x \in E, \quad \forall n \geqslant 1, \qquad \frac{1}{a_n} \leqslant g_n(x) \leqslant a_n \tag{1.3}$$

Using the fact that

$$\psi_n\left(\frac{1}{N} \sum_{i=1}^{N} \delta_{x^i}\right) = \sum_{i=1}^{N} \frac{g_n(x^i)}{\sum_{j=1}^{N} g_n(x^j)}\, \delta_{x^i}$$

we see that the resulting motion of the particles is decomposed into two separate mechanisms

$$\xi_{n-1} = (\xi_{n-1}^1, \ldots, \xi_{n-1}^N) \to \hat{\xi}_{n-1} = (\hat{\xi}_{n-1}^1, \ldots, \hat{\xi}_{n-1}^N) \to \xi_n = (\xi_n^1, \ldots, \xi_n^N)$$

These two transitions can be modelled as follows:

Initial particle system:

$$P_y(\xi_0 \in dx) = \prod_{p=1}^{N} \pi_0(dx^p)$$

Selection/Updating:

$$P_y(\hat{\xi}_{n-1} \in dx \mid \xi_{n-1} = z) = \prod_{p=1}^{N} \sum_{i=1}^{N} \frac{g_n(y_n - h_n(z^i))}{\sum_{j=1}^{N} g_n(y_n - h_n(z^j))} \delta_{z^i}(dx^p)$$

Mutation/Prediction:

$$P_y(\xi_n \in dz \mid \hat{\xi}_{n-1} = x) = \prod_{p=1}^{N} K_n(x^p, dz^p) \tag{1.4}$$

Thus, we see that the particles move according to the following rules. In the selection transition, one updates the positions in accordance with the fitness functions $\{g_n; n \geqslant 1\}$ and the current configuration. More precisely, at each time $n \geqslant 1$, each particle examines the system of particles $\xi_{n-1} = (\xi_{n-1}^1, \ldots, \xi_{n-1}^N)$ and chooses randomly a site ξ_{n-1}^i, $1 \leqslant i \leqslant N$, with a probability which depends on the entire configuration ξ_{n-1} and given by

$$\frac{g_n(\xi_{n-1}^i)}{\sum_{j=1}^{N} g_n(\xi_{n-1}^j)}$$

This mechanism is called the Selection/Updating transition as the particles are selected for reproduction, the most fit individuals being more likely to be selected. In other words, this transition allows particles to give birth to some particles at the expense of light particles which die. The second mechanism is called Mutation/Prediction since at this step each particle evolves randomly according to a given transition probability kernel.

The preceding scheme is clearly a system of interacting particles undergoing adaptation in a time nonhomogeneous environment represented by the fitness functions $\{g_n; n \geqslant 1\}$. Roughly speaking the natural idea is to approximate the two step transitions of the system (1.1)

$$\pi_{n-1} \xrightarrow{\text{Updating}} \hat{\pi}_{n-1} \stackrel{\text{def}}{=} \psi_{n-1}(\pi_{n-1}) \xrightarrow{\text{Prediction}} \pi_n = \hat{\pi}_{n-1} K_{n-1}$$

by a two step Markov chain taking values in the set of finitely discrete probability measures with atoms of size some integer multiple of $1/N$. Namely,

$$\pi_{n-1}^N = \frac{1}{N}\sum_{i=1}^N \delta_{\xi_{n-1}^i} \xrightarrow{\text{Selection}} \hat{\pi}_{n-1}^N = \frac{1}{N}\sum_{i=1}^N \delta_{\hat{\xi}_{n-1}^i} \xrightarrow{\text{Mutation}} \pi_n = \frac{1}{N}\sum_{i=1}^N \delta_{\xi_n^i}$$

1.4. Statement of the Main Results

Given a collection $\mathscr{F}$ of measurable functions $f: E \to \mathbb{R}$, the particle density profiles π_n^N at time n induce a map from $\mathscr{F}$ to $\mathbb{R}$ given by

$$\pi_n^N: f \in \mathscr{F} \to \pi_n^N(f) \in \mathbb{R}$$

The $\mathscr{F}$-indexed collection $\{\pi_n^N(f); f \in \mathscr{F}\}$ is usually called the $\mathscr{F}$-empirical process associated to the empirical random measures π_n^N. The semi-metric commonly used in such a context is the Zolotarev semi-norm defined by

$$\|\mu - \nu\|_{\mathscr{F}} = \sup\{|\mu(f) - \nu(f)|; f \in \mathscr{F}\}, \qquad \forall \mu, \nu \in \mathbf{M}_1(E)$$

[see for instance Rachev[32]]. In order to control the behavior of the supremum $\|\pi_n^N - \pi_n\|_{\mathscr{F}}$ as $N \to \infty$, we will impose conditions on the class $\mathscr{F}$ that are classicaly used in the statistical theory of empirical processes for independent samples. To avoid technical measurability conditions, and in order not to obscure the main ideas, we will always assume the class $\mathscr{F}$ to be countable and uniformly bounded. Our conclusions also hold under appropriate separability assumptions on the empirical process. We do not enter these questions here [see van der Vaart and Wellner[38]]. The Glivenko–Cantelli and Donsker theorems are uniform versions of the law of large numbers and the central limit theorem for the empirical measures. In the classical theory of independent random variables, these properties are usually shown to hold under entropy conditions on the class $\mathscr{F}$. Namely, to measure the size of a given class $\mathscr{F}$, one considers the covering numbers $N(\varepsilon, \mathscr{F}, L_p(\mu))$ defined as the minimal number of $L_p(\mu)$-balls of radius $\varepsilon > 0$ needed to cover $\mathscr{F}$. With respect to the classical theory, we will need assumptions on these covering numbers uniformly over all probability measures μ. Classically also, this supremum can be taken over all discrete probability measures. Since we are dealing with interacting particle schemes, we however need to strengthen the assumption and take the corresponding supremum over all probability measures. Several examples of classes of functions satisfying the foregoing uniform entropy conditions are discussed [see van der Vaart and Wellner[38]]. Denote thus by

$\mathcal{N}(\varepsilon, \mathcal{F})$, $\varepsilon > 0$, and by $I(\mathcal{F})$ the uniform covering numbers and entropy integral given by

$$\mathcal{N}(\varepsilon, \mathcal{F}) = \sup\{N(\varepsilon, \mathcal{F}, L_2(\mu)); \mu \in \mathbf{M}_1(E)\}$$

$$I(\mathcal{F}) = \int_0^1 \sqrt{\log \mathcal{N}(\varepsilon, \mathcal{F})}\, d\varepsilon$$

Our results will then take the following form.

Theorem 1 [Glivenko–Cantelli]. If $\mathcal{N}(\varepsilon, \mathcal{F}) < \infty$ for any $\varepsilon > 0$, then, for any time $n \geqslant 0$,

$$\lim_{N \to \infty} \|\pi_n^N - \pi_n\|_{\mathcal{F}} = 0 \qquad P.\text{ a.s.}$$

If $I(\mathcal{F}) < \infty$, we will prove that the $\mathcal{F}$-indexed process W_n^N is asymptotically tight. Together with the central limit theorem for finite dimensional distributions described by Del Moral and Guionnet,[14] this leads to a uniform version in the Banach space $l^\infty(\mathcal{F})$ of all bounded functions on $\mathcal{F}$.

Theorem 2 [Donsker]. Under some regularity conditions on the transition kernels K_n, if $I(\mathcal{F}) < \infty$, then, for any time $n \geqslant 0$, the empirical process

$$W_n^N : f \in \mathcal{F} \to W_n^N(f) = \sqrt{N}(\pi_n^N(f) - \pi_n(f))$$

converges in law in $l^\infty(\mathcal{F})$ to a centered Gaussian process $\{W_n(f); f \in \mathcal{F}\}$.

The proofs of Theorems 1 and 2 are given in Section 2. Our method of proof is essentially based on a precise study of the dynamical structure of the limiting measure valued process (1.2) and on properties of covering numbers. For reasons which will appear clearly later on, this approach provides a natural and simple way to extend several results of the classical theory of empirical processes to genetic-type interacting particle systems. For these reasons, we present at the very beginning of Section 2 some key lemmas that will be used in extending these results.

Section 3 is concerned with further uniform results. We first establish a precise exponential bound for classes of functions $\mathcal{F}$ with polynomial covering numbers, such as the well-known Vapnik–Cervonenkis classes. More precisely we will prove the following extension of a theorem due to Talagrand[37] in the independent case.

Theorem 3. Let $\mathcal{F}$ be a countable class of measurable functions $f: E \to [0, 1]$ satisfying

$$\mathcal{N}(\varepsilon, \mathcal{F}) \leqslant \left(\frac{C}{\varepsilon}\right)^V \qquad \text{for every} \quad 0 < \varepsilon < C$$

for some constants C and V. Then, for any $n \geqslant 0$, $\delta > 0$ and $N \geqslant 1$, we have

$$P(\| W_n^N \|_{\mathcal{F}} > \delta \sigma_n) \leqslant (n+1) \left(\frac{D\delta}{\sqrt{V}}\right)^V e^{-2\delta^2}$$

where D, resp. σ_n, is a constant that only depends on C, resp. on the fitness functions $\{g_n; n \geqslant 1\}$.

In the second part of Section 3, we present a uniform convergence result with respect to time under some additional conditions on the limiting system (1.2).

Theorem 4. When the class of function $\mathcal{F}$ is sufficiently regular and under certain conditions on the mappings $\{\phi_n; n \geqslant 1\}$, there exists a convergence rate $\alpha > 0$ such that

$$\sup_{n \geqslant 0} E(\| \pi_n^N - \pi_n \|_{\mathcal{F}}) \leqslant \frac{C}{N^{\alpha/2}} I(\mathcal{F})$$

for some constant C that does not depend on the time parameter $n \geqslant 0$.

Application of these results to nonlinear filtering problems are presented in the last section.

2. GLIVENKO–CANTELLI AND DONSKER THEOREMS

Before turning to the proof of Theorems 1 and 2, we present three simple lemmas on the dynamical structure of (1.2) and the covering numbers. These properties will also be useful in the further developments of Section 3.

Let us introduce before some additional notations. Denote by $\{\phi_{n|p}; 0 \leqslant p \leqslant n\}$ the composite mappings

$$\phi_{n|p} = \phi_n \circ \phi_{n-1} \circ \cdots \circ \phi_{p+1}, \qquad 0 \leqslant p \leqslant n$$

(with the convention $\phi_{n|n} = Id$). A clear backward induction on the parameter p shows that the composite mappings $\phi_{n|p}$ have the same form as the one step mappings $\{\phi_n; n \geqslant 1\}$. This is the content of the next lemma.

Lemma 1. For any $0 \leqslant p \leqslant n$, we have

$$\phi_{n|p}(\mu)(f) = \frac{\mu(g_{n|p} K_{n|p}(f))}{\mu(g_{n|p})}$$

for all $f \in \mathscr{C}_b(E)$ where

$$K_{n|p-1}(f) = \frac{K_p(g_{n|p} K_{n|p}(f))}{K_p(g_{n|p})}, \qquad g_{n|p-1} = g_p K_p(g_{n|p}) \tag{2.1}$$

with the conventions $g_{n|n} = 1$ and $K_{n|n} = Id$.

Under our assumptions and using these notations we also notice that

$$\forall x \in E, \qquad \forall 0 \leqslant p \leqslant n, \qquad \frac{1}{a_{n|p}} \leqslant g_{n|p}(x) \leqslant a_{n|p} \tag{2.2}$$

where

$$a_{n|p} = \prod_{q=p+1}^{n} a_q, \qquad 0 \leqslant p \leqslant n$$

Lemma 2. For any $0 \leqslant p \leqslant n$, any $\mu, \nu \in \mathbf{M}_1(E)$ and any function f

$$\phi_{n|p}(\mu)(f) - \phi_{n|p}(\nu)(f)$$
$$= \frac{1}{\nu(g_{n|p})} \left[(\mu(f_{n|p}) - \nu(f_{n|p})) + \phi_{n|p}(\mu)(f)(\nu(g_{n|p}) - \mu(g_{n|p})) \right] \tag{2.3}$$

where

$$f_{n|p} \stackrel{\text{def}}{=} g_{n|p} K_{n|p}(f)$$

If, for any $g: E \to \mathbb{R}$ such that $\|g\| \leqslant 1$ and for any Markov kernel K on E we define

$$g \cdot K\mathscr{F} = \{ gK(f);\ f \in \mathscr{F} \}$$

then we have

Lemma 3. For any $p \geqslant 1$, $\varepsilon > 0$ and $\mu \in \mathbf{M}_1(E)$,

$$N(\varepsilon, g \cdot K\mathscr{F}, L_p(\mu)) \leqslant N(\varepsilon, \mathscr{F}, L_p(\mu K))$$

and therefore,

$$\mathcal{N}(\varepsilon, g \cdot K\mathcal{F}) \leqslant \mathcal{N}(\varepsilon, \mathcal{F}), \qquad I(g \cdot K\mathcal{F}) \leqslant I(\mathcal{F})$$

Lemma 3 follows from the fact that $N(\varepsilon, g \cdot \mathcal{F}, L_p(\mu)) \leqslant N(\varepsilon, \mathcal{F}, L_p(\mu))$ and $N(\varepsilon, K\mathcal{F}, L_p(\mu)) \leqslant N(\varepsilon, \mathcal{F}, L_p(\mu K))$. The first assertion is trivial. To establish the second inequality, simply note that, for every function f, $|K(f)|^p \leqslant K(|f|^p)$ and go back to the definition of the covering numbers.

2.1. Glivenko–Cantelli Theorem

To establish a Glivenko–Cantelli property in our setting, we make use of the following basic decomposition: for any $f \in \mathcal{F}$,

$$\pi_n^N(f) - \pi_n(f) = \sum_{p=0}^{n} [\phi_{n|p}(\pi_p^N)(f) - \phi_{n|p}(\phi_p(\pi_{p-1}^N))(f)] \tag{2.4}$$

with the convention $\phi_0(\pi_{-1}^N) = \pi_0$. There is no loss of generality in assuming that $1 \in \mathcal{F}$. Then, by Lemma 2, we get

$$\|\phi_{n|p}(\mu) - \phi_{n|p}(\nu)\|_{\mathcal{F}} \leqslant 2a_{n|p}^2 \|\mu - \nu\|_{\mathcal{F}_{n|p}} \tag{2.5}$$

where $\mathcal{F}_{n|p} = \bar{g}_{n|p} \cdot K_{n|p}\mathcal{F}$ and $\bar{g}_{n|p} = a_{n|p}^{-1} g_{n|p}$ so that $\|\bar{g}_{n|p}\| \leqslant 1$). It easily follows that, for every $\varepsilon > 0$,

$$\begin{aligned} &P(\|\pi_n^N - \pi_n\|_{\mathcal{F}} > \varepsilon) \\ &\qquad \leqslant (n+1) \sup_{0 \leqslant p \leqslant n} P\left(\|\phi_{n|p}(\pi_p^N) - \phi_{n|p}(\phi_p(\pi_{p-1}^N))\|_{\mathcal{F}} > \frac{\varepsilon}{n+1} \right) \end{aligned}$$

Using (2.5), this implies that

$$P(\|\pi_n^N - \pi_n\|_{\mathcal{F}} > \varepsilon) \leqslant (n+1) \sup_{0 \leqslant p \leqslant n} P\left(\|\phi_p^N - \phi_p(\pi_{p-1}^N)\|_{\alpha} > \frac{\varepsilon}{\sigma_n} \right) \tag{2.6}$$

where $\sigma_n = 2(n+1)\, a_{n|0}^2$.

Our main assumption in this section will be

(H1) $\forall \varepsilon > 0,\ \mathcal{N}(\varepsilon, \mathcal{F}) < \infty$

The Glivenko–Cantelli theorem may then be stated as follows.

Theorem 5. Assume that $\mathcal{F}$ is a countable collection of functions f such that $\|f\| \leqslant 1$ and (H1) holds. Then, for any time $n \geqslant 0$, $\|\pi_n^N - \pi_n\|_{\mathcal{F}}$ converges almost surely to 0 as $N \to \infty$.

Proof. It is based on the following standard lemma in the theory of empirical processes.

Lemma 4. Let $\{X^{i,N}; 1 \leqslant i \leqslant N\}$ be independent random variables with common law $P^{(N)}$ and let $\mathscr{F}$ be a countable collection of functions f such that $\|f\| \leqslant 1$. Then, for any $\varepsilon > 0$ and $\sqrt{N} \geqslant 4\varepsilon^{-1}$ we have that

$$P(\|m^N(X) - P^{(N)})\|_{\mathscr{F}} > 8\varepsilon) \leqslant 8\mathscr{N}(\varepsilon, \mathscr{F})\, e^{-N\varepsilon^2/2}$$

where $m^N(X) = (1/N) \sum_{i=1}^N \delta_{X^{i,N}}$.

Before turning to the sketch of the proof of this lemma, let us show how it implies the theorem. n is fixed throughout the argument. Using (2.6) and Lemma 4 one easily gets that, for $\sqrt{N} \geqslant 4\varepsilon_n^{-1}$ where $\varepsilon_n = \varepsilon/8\sigma_n$,

$$P(\|\pi_n^N - \pi_n\|_{\mathscr{F}} > \varepsilon) \leqslant 8(n+1)\, e^{-N\varepsilon_n^2/2} \sup_{0 \leqslant p \leqslant n} \mathscr{N}(\varepsilon_n, \mathscr{F}_{n|p})$$

Furthermore by definition of the class $\mathscr{F}_{n|p}$ and with the help of Lemma 3, we know that, for each $\varepsilon > 0$, and $0 \leqslant p \leqslant n$,

$$\mathscr{N}(\varepsilon, \mathscr{F}_{n|p}) \leqslant \mathscr{N}(\varepsilon, \mathscr{F})$$

Therefore,

$$P(\|\pi_n^N - \pi_n\|_{\mathscr{F}} > \varepsilon) \leqslant 8(n+1)\, \mathscr{N}(\varepsilon_n, \mathscr{F})\, e^{-N\varepsilon_n^2/2}$$

as soon as $\sqrt{N} \geqslant 4\varepsilon_n^{-1}$. The end of proof of Theorem 5 is an immediate consequence of the Borel–Cantelli lemma.

Proof of Lemma 4. Using the classical symmetrization inequalities [see van der Vaart and Wellner,[38] Lem. 2.3.7 or Pollard,[31] pp. 14–15] for any $\varepsilon > 0$ and $\sqrt{N} \geqslant 4\varepsilon^{-1}$,

$$P(\|m^N(X) - P^{(N)}\|_{\mathscr{F}} > \varepsilon) \leqslant 4P\left(\|m_\varepsilon^N(X)\|_{\mathscr{F}} > \frac{\varepsilon}{4}\right) \tag{2.7}$$

where $m_\varepsilon^N(X)$ denotes the signed measure

$$m_\varepsilon^N(X) = \frac{1}{N} \sum_{i=1}^N \varepsilon_i\, \delta_{X^{i,N}}$$

and $\{\varepsilon_1,..., e_N\}$ are symmetric Bernoulli random variables, independent of the $X^{i,N}$'s. Conditionally on the $X^{i,N}$'s, and by definition of the covering numbers, we easily get by a standard argument that

$$P(\|m_\varepsilon^N(X)\|_{\mathscr{F}} > \delta \mid X) \leqslant N(\delta/2, \mathscr{F}, L^1(m^N(X))) \sup_{f \in \mathscr{F}} P(|m_\varepsilon^N(X)(f)| > \delta/2 \mid X) \tag{2.8}$$

Indeed, let

$$\{f_m; 1 \leqslant m \leqslant N(\delta/2, \mathscr{F}, L^1(m^N(X)))\}$$

be a $(\delta/2)$-coverage of $\mathscr{F}$ for the $L^1(m^N(X))$-norm. Then,

$$P(\|m_\varepsilon^N(X)\|_{\mathscr{F}} > \delta \mid X) \leqslant P(\sup_m |m_\varepsilon^N(X)(f_m)| > \delta/2 \mid X)$$

Therefore,

$$P(\|m_\varepsilon^N(X)\|_{\mathscr{F}} > \delta \mid X) \leqslant N(\delta/2, \mathscr{F}, L^1(m^N(X))) \sup_m P(|m_\varepsilon^N(X)(f_m)| > \delta/2 \mid X)$$

By Hoeffding's inequality [see for instance van der Vaart and Wellner,[38] Lem. 2.2.7], for any $f \in \mathscr{F}$ and $\delta > 0$,

$$P(|m_\varepsilon^N(X)(f)| > \delta/2 \mid X) \leqslant 2e^{-N\delta^2/8}$$

As a consequence, we see that the conditional probability

$$P\left(\|m_\varepsilon^N(X)\|_{\mathscr{F}} > \frac{\varepsilon}{4} \mid X\right)$$

is bounded above by

$$2N(\varepsilon/8, \mathscr{F}, L_1(m^N(X)))\, e^{-N\varepsilon^2/128} \leqslant 2\mathscr{N}(\varepsilon/8, \mathscr{F})\, e^{-N\varepsilon^2/128}$$

From (2.7), it follows that

$$P(\|m^N(X) - P^{(N)}\|_{\mathscr{F}} > \varepsilon) \leqslant 8\mathscr{N}(\varepsilon/8, \mathscr{F})\, e^{-N\varepsilon^2/128}$$

as soon as $\sqrt{N} \geqslant 4\varepsilon^{-1}$ which is the result. The proof of Lemma 4 is complete. $\square$

2.2. Donsker's Theorem

After the Glivenko–Cantelli theorem, Donsker's theorem is the second most important result on the convergence of empirical processes. Theorem 2 that we will establish in this section, may be seen as a uniform version of the central limit theorem presented by Del Moral and Guionnet.[14] More precisely, it was shown there that the marginals of the $\mathscr{F}$-indexed process $\{W_n^N(f); f \in \mathscr{F}\}$ converge weakly to the marginal of a $\mathscr{F}$-indexed centered Gaussian process $\{W_n(f); f \in \mathscr{F}\}$. Now, weak convergence in $l^\infty(\mathscr{F})$ can be characterized as the convergence of the marginals together with the asymptotic tightness of the process $\{W_n^N(f); f \in \mathscr{F}\}$. In view of these observations, the proof of Theorem 2 is naturally decomposed into two steps. In the first step, we establish the asymptotic tightness using the standard entropy condition

(H2) $I(\mathscr{F}) < \infty$

We then make use of the results by Del Moral and Guionnet,[14] to identify the limiting Gaussian process. To this task, and as in Ref. 15, we need to impose some further conditions on the kernels K_n. We will namely assume that

(H3) K_n is Feller for each $n \geqslant 1$ and such that there exists, for any time $n \geqslant 1$, a reference probability measure $\lambda_n \in \mathbf{M}_1(E)$ and a $\mathbf{B}(E)$-measurable non-negative function φ_n such that

$$K_n(x, \cdot) \sim \lambda_n$$

for every $x \in E$ with

$$\left| \log \frac{dK_n(x, \cdot)}{d\lambda_n} \right| \leqslant \varphi_n \qquad \text{and} \qquad \int e^{r\varphi_n} \, d\lambda_n < \infty \tag{2.9}$$

for every $r \geqslant 1$.

Theorem 6. Assume that $\mathscr{F}$ is a countable class of functions such that $\|f\| \leqslant 1$ for any $f \in \mathscr{F}$ and satisfying (H2) and (H3). Then, for any $n \geqslant 0$, $\{W_n^N(f); f \in \mathscr{F}\}$ converges weakly in $l^\infty(\mathscr{F})$ as $N \to \infty$ to a centered Gaussian process $\{W_n(f); f \in \mathscr{F}\}$ with covariance function

$$\begin{aligned} &E(W_n(f)\, W_n(h)) \\ &\quad = \sum_{p=0}^{n} \int \left(\frac{g_{n|p}}{\pi_p(g_{n|p})} \right)^2 (K_{n|p}(f) - \pi_n(f))(K_{n|p}(h) - \pi_n(h)) \, d\pi_p \end{aligned}$$

Remark 1. If the transition probability kernels $\{K_n; n \geqslant 1\}$ are trivial, in the sense that, for every n,

$$K_n(x, dz) = \mu_n(dz), \qquad \mu_n \in \mathbf{M}_1(E)$$

then one can readily check that

$$\pi_n = \mu_n \qquad \text{and} \qquad K_{n|p}(x, dz) = \mu_n(dz), \qquad 0 \leqslant p \leqslant n$$

In this particular situation $\{W_n(f); f \in L_2(\pi_n)\}$ is the classical μ_n-Brownian bridge. Namely, W_n is the centered Gaussian process with covariance

$$E(W_n(f)\, W_n(h)) = \mu_n((f - \mu_n f)(h - \mu_n h))$$

As announced, the proof of Theorem 6 will be a consequence of the following two lemmas. Note that the entropy condition $I(\mathscr{F}) < \infty$ is only used in the proof of the asymptotic tightness, while the regularity condition (H3) is needed in. identifying the limiting Gaussian process. Again, n is fixed throughout the proof of Theorem 6.

Lemma 5. If $\mathscr{F}$ is a countable collection of functions f such that $\|f\| \leqslant 1$ and (H2) holds then the $\mathscr{F}$-indexed process $\{W_n^N(f); f \in \mathscr{F}\}$ is asymptotically tight.

Lemma 6. Under (H3), the marginals of $\{W_n^N(f); f \in L_2(\pi_n)\}$ converge weakly to the marginals of a centered Gaussian process $\{W_n(f); f \in L_2(\pi_n)\}$ with covariance function

$$E(W_n(f)\, W_n(h)) = \sum_{p=0}^{n} \int \left(\frac{g_{n|p}}{\pi_p(g_{n|p})} \right)^2 (K_{n|p}(f) - \pi_n(f))(K_{n|p}(h) - \pi_n(h))\, d\pi_p$$

Since the collection of functions $\{g_{n|p}; 0 \leqslant p \leqslant n\}$ satisfies (2.2), using the Cauchy–Schwarz inequality and Lemma 1, we have that

$$E(|W_n(f) - W_n(h)|^2) \leqslant C_n \|f - h\|^2_{L_2(\pi_n)}$$

for some constant $C_n < \infty$ and all $f, h \in L_2(\pi_n)$. In particular, to prove the asymptotic tightness in Lemma 5, it will be enough to establish the asymptotic equicontinuity in probability of $\{W_n^N(f); f \in \mathscr{F}\}$ with respect to the semi-norm on $\mathscr{F}$ given by

$$f \in \mathscr{F} \to \pi_n((f - \pi_n(f))^2)^{1/2}$$

[see for instance van der Vaart and Wellner,[38] Chap. 1.5; Ex. 1.5.10, p. 40].

Proof of Lemma 5. Under (H2), the class $\mathscr{F}$ is totally bounded in $L_2(\pi_n)$ for any $n \geqslant 0$. According to the preceeding comment, it will be enough to show that

$$\lim_{N \to \infty} E(\| W_n^N \|_{\mathscr{F}^{(n)}(\delta_N)}) = 0 \tag{2.10}$$

for all sequences $\delta_N \downarrow 0$ where, for any $\delta > 0$ (possibly infinite),

$$\mathscr{F}^{(n)}(\delta) \stackrel{\text{def}}{=} \{ f - h;\ f, h \in \mathscr{F} : \| f - h \|_{L_2(\pi_n)} < \delta \}$$

To this task, we use the decompositions (2.4) and (2.3) to get that

$$\begin{aligned} \| \pi_n^N - \pi_n \|_{\mathscr{F}^{(n)}(\delta)} \leqslant \sum_{p=0}^{n} a_{n|p}^2 [& \| \pi_p^N - \phi_p(\pi_{p-1}^N) \|_{\mathscr{F}_{n|p}(\delta)} \\ & + \| \pi_p^N \|_{\mathscr{F}_{n|p}(\delta)} \, | \pi_p^N(\bar{g}_{n|p}) - \phi_p(\pi_{p-1}^N)(\bar{g}_{n|p}) |] \end{aligned}$$

where

$$\mathscr{F}_{n|p}(\delta) = \{ \bar{g}_{n|p} K_{n|p} f;\ f \in \mathscr{F}^{(n)}(\delta) \}$$

$\bar{g}_{n|p} = (1/a_{n|p})\, g_{n|p}$ and $a_{n|p} = \prod_{q=p+1}^{n} a_q$. Since for any $0 \leqslant p \leqslant n$,

$$E([\pi_p^N(\bar{g}_{n|p}) - \phi_p(\pi_{p-1}^N)(\bar{g}_{n|p})]^2) \leqslant \frac{1}{N}$$

to prove (2.10) it suffices to check that for any $0 \leqslant p \leqslant n$ and $\delta_N \downarrow 0$

$$(1) \quad \lim_{N \to \infty} E(\sqrt{N}\, \| \pi_p^N - \phi_p(\pi_{p-1}^N) \|_{\mathscr{F}_{n|p}(\delta_N)}) = 0$$

$$(2) \quad \lim_{N \to \infty} E(\| \pi_p^N \|_{\mathscr{F}^2_{n|p}(\delta_N)}) = 0$$

where

$$\mathscr{F}^2_{n|p}(\delta) = \{ f^2;\ f \in \mathscr{F}_{n|p}(\delta) \}$$

Let us prove (1). By the symmetrization inequalities, for any N,

$$E(\| \pi_p^N - \phi_p(\pi_{p-1}^N) \|_{\mathscr{F}_{n|p}(\delta_N)}) \leqslant 2 E(\| m_\varepsilon^N(\xi_p) \|_{\mathscr{F}_{n|p}(\delta_N)})$$

where, as in Section 2.1,

$$m_\varepsilon^N(\xi_p) = \frac{1}{N} \sum_{i=1}^{N} \varepsilon_i \, \delta_{\xi_p^i}$$

Fix $\xi_p = (\xi_p^1, ..., \xi_p^N)$. By Hoeffding's inequality [cf. van der Vaart and Wellner[38], Lem. 2.2.7], the process

$$f \to \frac{1}{\sqrt{N}} \sum_{i=1}^{N} f(\xi_p^i)\, \varepsilon_i = \sqrt{N}\, m_\varepsilon^N(\xi_p)(f)$$

is sub-Gaussian with respect to the norm $\|\cdot\|_{L_2(\pi_p^N)}$. Namely, for any $f, h \in \mathscr{F}_{n|p}(\delta_N)$ and $\gamma > 0$,

$$P(|\sqrt{N}(m_\varepsilon^N(\xi_p)(f) - m_\varepsilon^N(\xi_p)(h))| > \gamma \mid \xi_p) \leqslant 2e^{-(1/2)\, \gamma^2/\|f-h\|^2_{L_2(\pi_p^N)}}$$

Using the maximal inequality for sub-Gaussian processes [cf. Ledoux and Talagrand,[26] and van der Vaart and Wellner[38]], we get the quenched inequality

$$\begin{aligned} &E(\|m_\varepsilon^N(\xi_p)\|_{\mathscr{F}_{n|p}(\delta_N)} \mid \xi_p) \\ &\quad \leqslant \frac{C}{\sqrt{N}} \int_0^{\theta_{n\,|\,p}(N)} \sqrt{\log N(\varepsilon, \mathscr{F}_{n|p}(\delta_N), M_2(\pi_p^N))}\, d\varepsilon \qquad (2.11) \end{aligned}$$

where

$$\theta_{n|p}(N) = \|\pi_p^N\|_{\mathscr{F}^2_{n|p}(\delta_N)}$$

On the other hand we clearly have that, for every $\varepsilon > 0$,

$$N(\varepsilon, \mathscr{F}_{n|p}, L_2(\pi_p^N)) \leqslant N(\varepsilon, \mathscr{F}_{n|p}(\infty), L_2(\pi_p^N)) \leqslant N^2(\varepsilon/2, \mathscr{F}_{n|p}, L_2(\pi_p^N))$$

where we recall that $\mathscr{F}_{n|p} = \bar{g}_{n|p} \cdot K_{n|p}\mathscr{F}$. Under our assumptions, it thus follows from Lemma 3 that

$$N(\varepsilon, \mathscr{F}_{n|p}(\delta), L_2(\pi_p^N)) \leqslant \mathscr{N}^2(\varepsilon/2, \mathscr{F})$$

Using (2.11), one concludes that, for every $N \geqslant 1$,

$$E(\|m_\varepsilon^N(\xi_p)\|_{\mathscr{F}_{n|p}(\delta_N)}) \leqslant \frac{\sqrt{2}\, C}{\sqrt{N}}\, E\left(\int_0^{\theta_{n\,|\,p}(N)} \sqrt{\log \mathscr{N}(\varepsilon/2, \mathscr{F})}\, d\varepsilon\right)$$

and therefore

$$E(\|\pi_p^N - \phi_p(\pi_{p-1}^N)\|_{\mathscr{F}_{n|p}(\delta_N)}) \leqslant \frac{2\sqrt{2}\,C}{\sqrt{N}}\, E\left(\int_0^{\theta_{n|p}(N)} \sqrt{\log \mathscr{N}(\varepsilon/2, \mathscr{F})}\, d\varepsilon\right)$$

By the dominated convergence theorem, to prove (1) it suffices to check that

$$\lim_{N\to\infty} \theta_{n|p}(N) = \lim_{N\to\infty} \|\pi_p^N\|_{\mathscr{F}^2_{n|p}(\delta_N)} = 0 \quad P.\ \text{a.s.} \tag{2.12}$$

We establish this property by proving that

$$\text{(a)} \quad \|\pi_p\|_{\mathscr{F}^2_{n|p}(\delta_N)} \leqslant \delta_N^2$$

$$\text{(b)} \quad \lim_{N\to\infty} \|\pi_p^N - \pi_p\|_{\mathscr{F}^2_{n|p}(\delta_N)} = 0 \quad P.\ \text{a.s.}$$

Let $f, h \in \mathscr{F}$ be chosen so that $\pi_n((f-h)^2) < \delta^2$ (i.e., $f-h \in \mathscr{F}^{n)}(\delta)$). Use Cauchy–Schwartz to see that

$$\pi_p([\bar{g}_{n|p} K_{n|p}(f-h)]^2) \leqslant \pi_p(\bar{g}^2_{n|p} K_{n|p}((f-h)^2)) \tag{2.13}$$

Since $0 \leqslant \bar{g}_{n|p} = a^{-1}_{n|p} g_{n|p} \leqslant 1$, the right-hand side of (2.13) is bounded above by

$$\frac{1}{a_{n|p}}\, \pi_p(g_{n|p} K_{n|p}((f-h)^2)) = \frac{\pi_p(g_{n|p})}{a_{n|p}}\, \pi_n((f-h)^2)$$

which is less that δ^2. This ends the proof of (a). To prove (b), first note that

$$\|\pi_p^N - \pi_p\|_{\mathscr{F}^2_{n|p}(\delta_N)} \leqslant \|\pi_p^N - \pi_p\|_{\mathscr{F}^2_{n|p}(\infty)}$$

Using Theorem 5 to prove (b) it certainly suffices to show that

$$\sup_\mu N(\varepsilon, \mathscr{F}^2_{n|p}(\infty), L_2(\mu)) < \infty$$

for every $\varepsilon > 0$. Since all functions in $\mathscr{F}$ have norm less than or equal to 1, for any f, h in $\mathscr{F}_{n|p}(\infty)$ and any $\mu \in \mathbf{M}_1(E)$,

$$\|f^2 - h^2\|_{L_2(\mu)} \leqslant 4\, \|f-h\|_{L_2(\mu)}$$

It follows that, for every $\varepsilon > 0$,

$$N(\varepsilon, \mathscr{F}^2_{n|p}(\infty), L_2(\mu)) \leqslant N(\varepsilon/4, F_{n|p}(\infty), L_2(\mu))$$

Since

$$N(\varepsilon, \mathscr{F}_{n|p}(\infty), L_2(\mu)) \leqslant N^2(\varepsilon/2, F_{n|p}, L_2(\mu))$$

one concludes, using Lemma 3 that

$$\sup_{\mu} N(\varepsilon, \mathscr{F}^2_{n|p}(\infty), L_2(\mu)) \leqslant \sup_{\mu} N^2(\varepsilon/8, \mathscr{F}, L_2(\mu))$$

This ends the proof of (b) and (1). In the same way, by dominated convergence, the proof of (2) is an immediate consequence of (2.12). This completes the proof of Lemma 5. □

We turn to the proof of Lemma 6. To this task, we need to recall some results presented in Del Moral and Guionnet.[14] Under (H3), for any $n \geqslant 1$, set

$$\forall (x, z) \in E^2, \qquad k_n(x, z) \stackrel{\text{def}}{=} \frac{dK_n(x, \cdot)}{d\lambda_n}(z)$$

From now on, the time parameter $n \geqslant 0$ is fixed. For any $x = (x_0,..., x_n)$ and $z = (z_0,..., z_n) \in E^{n+1}$ set,

$$q_{[0,n]}(x, z) = \sum_{m=1}^{n} q_m(x, z)$$

with

$$q_m(x, z) = \frac{g_m(z_{m-1})\, k_m(z_{m-1}, x_m)}{\int_E g_m(u)\, k_m(u, x_m)\, \pi_{m-1}(du)}$$

$$a_{[0,n]}(x, z) = q_{[0,n]}(x, z) - \int_{\Sigma_n} q_{[0,n]}(x', z)\, \pi_{[0,n]}(dx')$$

$$\pi_{[0,n]} = \pi_0 \otimes \cdots \otimes \pi_n$$

Observe now that from the exponential moment condition (2.9) of (H3), $q_{[0,n]} \in L^2(\pi_{[0,n]} \otimes \pi_{[0,n]})$. It follows that $a_{[0,n]} \in L^2(\pi_{[0,n]} \otimes \pi_{[0,n]})$ and therefore the integral operator A_n given by

$$A_n \varphi(x) = \int a_{[0,n]}(z, x)\, \varphi(z)\, \pi_{[0,n]}(dz), \qquad \varphi \in L^2(\pi_{[0,n]})$$

is Hilbert–Schmidt on $L^2(\pi_{[0,n]})$. Under (H3), it is proved in Del Moral and Guionnet[14] that the integral operator $I-A_n$ is invertible and that the random field

$$\varphi \in L^2(\pi_{[0,n]}) \to W^N_{[0,n]}(\varphi)=\sqrt{N}\left(\frac{1}{N}\sum_{i=1}^{N}\varphi(\xi^i_0,..., \xi^i_n)-\pi_{[0,n]}(\varphi)\right)$$

converges as $N\to\infty$, in the sense of convergence of finite dimensional distributions, to a centered Gaussian field $\{W_{[0,n]}(\varphi);\ \varphi\in L^2(\pi_{[0,n]})\}$ with covariance

$$\begin{aligned}&E(W_{[0,n]}(\varphi_1)\,W_{[0,n]}(\varphi_2))\\&\quad=((I-A_n)^{-1}(\varphi_1-\pi_{[0,n]}(\varphi_1)), (I-A_n)^{-1}(\varphi_2-\pi_{[0,n]}(\varphi_2)))_{L^2(\pi_{[0,n]})}\end{aligned}$$

for $\varphi_1, \varphi_2\in L^2(\pi_{[0,n]})$.

From these observation it follows that the process

$$f\in L_2(\pi_n)\to W^N_n(f)=\sqrt{N}(\pi^N_n(f)-\pi_n(f))$$

converges in the sense of convergences of finite dimensional distributions and as $N\to\infty$ to a centered Gaussian field $\{W_n(\varphi);\ f\in L^2(\pi_n)\}$ satisfying

$$\begin{aligned}&E(W_n(f)\,W_n(h))\\&\quad=((I-A_n)^{-1}(f-\pi_n(f))^{\otimes 1}, (I-A_n)^{-1}(h-\pi_n(h))^{\otimes 1})_{L^2(\pi_n)}\end{aligned}$$

for any $f, h\in L^2(\pi_n)$, where, for all $f\in L^2(\pi_n)$,

$$f^{\otimes 1}\stackrel{\text{def}}{=}\underbrace{1\otimes\cdots\otimes 1}_{(n-1)\text{ times}}\otimes f$$

In the next lemma, we identify the preceding covariance function and thus establish in this way Lemma 6.

Lemma 7. For every $f\in L^2(\pi_n)$, and every $z_0,..., z_n\in E$,

$$(I-A_n)^{-1}(f-\pi_n(f))^{\otimes 1}(z_0,..., z_n)=\sum_{p=0}^{n}\tilde{f}_{n|p}(z_p)$$

where the functions $\tilde{f}_{n|p}$, $0 \leqslant p \leqslant n$, are given by

$$\tilde{f}_{n|p} = \frac{g_{n|p}}{\pi_p(g_{n|p})} (K_{n|p}(f) - \pi_n(f)), \qquad 0 \leqslant p \leqslant n$$

with the convention $g_{n|n} = 1$ and $K_{n|n} = Id$.

Proof. Using Lemma 1, we first note that

$$\begin{aligned}\tilde{f}_{n|p-1} &= \frac{g_p K_p(g_{n|p})}{\pi_{p-1}(g_{n|p-1})} \left(\frac{K_p(g_{n|p} K_{n|p}(f))}{K_p(g_{n|p})} - \pi_n(f) \right) \\ &= \frac{g_p}{\pi_{p-1}(g_{n|p-1})} (K_p(g_{n|p} K_{n|p}(f)) - K_p g_{n|p} \pi_n(f))\end{aligned}$$

and therefore

$$\begin{aligned}\tilde{f}_{n|p-1} &= \frac{g_p \pi_p(g_{n|p})}{\pi_{p-1}(g_{n|p-1})} K_p \left(\frac{g_{n|p}}{\pi_p(g_{n|p})} (K_{n|p}(f) - \pi_n(f)) \right) \\ &= g_p \frac{\pi_p(g_{n|p})}{\pi_{p-1}(g_{n|p-1})} K_p(\tilde{f}_{n|p})\end{aligned}$$

Then, using the fact that

$$\pi_p(g_{n|p}) = \frac{\pi_{p-1}(g_p K_p(g_{n|p}))}{\pi_{p-1}(g_p)} = \frac{\pi_{p-1}(g_{n|p-1})}{\pi_{p-1}(g_p)}$$

one easily gets the backward recursion equations

$$\tilde{f}_{n|p-1} = \frac{g_p}{\pi_{p-1}(g_p)} K_p(\tilde{f}_{n|p}), \qquad 1 \leqslant p \leqslant n \tag{2.14}$$

By definition of A_n, for any $\varphi \in L^2(\pi_{[0,n]})$, we have that

$$A_n \varphi(z_0,..., z_n) = \sum_{m=1}^{n} \int \varphi(x_0,..., x_n)\, a_m(x_m, z_{m-1})\, \pi_{[0,n]}(dx_0,..., dx_n)$$

where, for every $1 \leqslant m \leqslant n$,

$$a_m(x_m, z_{m-1}) = \frac{g_m(z_{m-1})\, k_m(z_{m-1}, x_m)}{\pi_{m-1}(g_m k_m(\cdot, x_m))} - \frac{g_m(z_{m-1})}{\pi_{m-1}(g_m)}$$

On the other hand, we observe that since

$$\pi_m(dx_m) = \frac{\pi_{m-1}(g_m k_m(\cdot, x_m))}{\pi_{m-1}(g_m)}\, \lambda_m(dx_m)$$

and

$$K_m(z_{m-1}, dx_m) = k_m(z_{m-1}, x_m)\, \lambda_m(dx_m)$$

we have that

$$\frac{g_m(z_{m-1})\, k_m(z_{m-1}, x_m)}{\pi_{m-1}(g_m k_m(\cdot, x_m))}\, \pi_m(dx_m) = \frac{g_m(z_{m-1})}{\pi_{m-1}(g_m)}\, K_m(z_{m-1}, dx_m)$$

Therefore

$$\begin{aligned}(I - A_n)\, &\varphi(z_0,..., z_n)\\ &= \varphi(z_0,..., z_n) + \sum_{m=1}^{n} \frac{g_m(z_{m-1})}{\pi_{m-1}(g_m)} \int \varphi(x_0,..., x_n)\, \tilde{\pi}^m_{[0,n]}(z_{m-1}; dx_0,..., dx_n)\end{aligned} \tag{2.15}$$

where

$$\begin{aligned}\tilde{\pi}^m_{[0,n]}&(z_{m-1}; dx_0,..., dx_n)\\ &= \pi_{[0,m-1]}(dx_0,..., dx_{m-1})(\pi_m(dx_m) - K_m(z_{m-1}, dx_m))\\ &\quad \times \pi_{[m+1,n]}(dx_{m+1},..., dx_n)\end{aligned}$$

Choose now φ given by

$$\varphi(z_0,..., z_n) = \sum_{p=0}^{n} \tilde{f}_{n|p}(z_p), \qquad z_0,..., z_n \in E$$

Then, we get

$$(I - A_n)\, \varphi(z_0,..., z_n) = \varphi(z_0,..., z_n) - \sum_{m=1}^{n} \frac{g_m(z_{m-1})}{\pi_{m-1}(g_m)}\, K_m(\tilde{f}_{n|m})(z_{m-1})$$

Finally, using the backward equation (2.14), one concludes that

$$\begin{aligned}(I - A_n)\, \varphi(z_0,..., z_n) &= \varphi(z_0,..., z_n) - \sum_{m=1}^{n} \tilde{f}_{n|m-1}(z_{m-1})\\ &= \sum_{m=0}^{n} \tilde{f}_{n|m}(z_m) - \sum_{m=0}^{n-1} \tilde{f}_{n|m}(z_m)\end{aligned}$$

so that the result follows from

$$(I - A_n)\,\varphi(z_0,..., z_n) = f_{n|n}(z_n) = f(z_n) - \pi_n(f) \qquad \square$$

The nature of the limiting Gaussian field $\{W_n(f); f \in L_2(\pi_n)\}$ is now clearly determined. More precisely, for every $f, h \in L_2(\pi_n)$, we get from the preceding that

$$\begin{aligned} &E(W_n(f)\, W_n(h)) \\ &\quad = \sum_{p=0}^{n} \int \left(\frac{g_{n|p}}{\pi_p(g_{n|p})}\right)^2 (K_{n|p}(f) - \pi_n(f))(K_{n|p}(h) - \pi_n(h))\, d\pi_p \end{aligned}$$

Lemma 6 is therefore established in this way. This completes the proof of Donsker's theorem.

3. FURTHER UNIFORM RESULTS

3.1. Exponential Bounds

In Section 2, we have shown that, under rather general assumptions, the Glivenko–Cantelli theorem hold for the genetic type interacting scheme (1.1). The proof of Theorem 5 also gives an exponential rate of convergence but this result is only valid for a number of particles larger than some value depending on the time parameter. In this section, we refine this exponential bound in the case of uniformly bounded classes $\mathscr{F}$ with polynomial covering numbers. More precisely we will use the following assumption

$$\text{(H4)} \quad \mathscr{N}(\varepsilon, \mathscr{F}) \leqslant \left(\frac{C}{\varepsilon}\right)^V \qquad \text{for every} \quad 0 < \varepsilon < C$$

for some constants C and V. Several examples of classes of functions satisfying this condition are discussed [see van der Vaart and Wellner(38)]. For instance Vapnik–Cervonenkis classes $\mathscr{F}$ of index $V(\mathscr{F})$ and envelope function $F = 1$ satisfy (H4) with $V = 2(V(\mathscr{F}) - 1)$ and a constant C that only depends on V.

Theorem 7. Let $\mathscr{F}$ be a countable class of measurable functions $f: E \to [0, 1]$ satisfying (H4) for some constants C and V. Then, for any $n \geqslant 0$, $\delta > 0$ and $N \geqslant 1$,

$$P(\|W_n^N\|_{\mathscr{F}} > \delta\sigma_n) \leqslant (n+1)\left(\frac{D\delta}{\sqrt{V}}\right)^V e^{-2\delta^2}$$

where D is a constant that only depends on C and $\sigma_n = 2(n+1)\prod_{p=1}^n a_p^2$.

Proof. We will use the decomposition (2.4). Using the same notations as in the beginning of Section 2 and in the proof of Theorem 5, we have

$$P\left(\|\phi_{n|p}(\pi_p^N)-\phi_{n|p-1}(\pi_{p-1}^N)\|_{\mathscr{F}}>\frac{\varepsilon}{n+1}\right)\leqslant P(\|\pi_p^N-\phi_p(\pi_{p-1}^N)\|_{\mathscr{F}_{n|p}}>\varepsilon_n) \tag{2.16}$$

where $\varepsilon_n=\varepsilon/\sigma_n$ and $\sigma_n=2(n+1)\,a_{n|0}^2$. Now, under the polynomial assumption (H4) on the covering numbers, it is convenient to note that the classes $\mathscr{F}_{n|p}$, $0\leqslant p\leqslant n$, also satisfy the assumptions of the theorem. Indeed, the class $\mathscr{F}_{n|p}$ is again a countable class of functions $f\colon E\to[0,1]$ and using Lemma 3 we also have, for every $\varepsilon>0$,

$$\mathscr{N}(\varepsilon,\mathscr{F}_{n|p})\leqslant\mathscr{N}(\varepsilon,\mathscr{F})\leqslant\left(\frac{C}{\varepsilon}\right)^V$$

We are now in position to apply the exponential bounds of Talagrand.[37] More precisely, by recalling that π_p^N is the empirical measure associated to N conditionally independent random variables with common law $\phi_p(\pi_{p-1}^N)$, and using the exponential bounds of Talagrand[37] [see also van der Vaart and Wellner[38]] we get

$$P(\|\pi_p^N-\phi_p(\pi_{p-1}^N)\|_{\mathscr{F}_{n|p}}>\varepsilon_n\mid\pi_{p-1}^N)\leqslant\left(\frac{D\sqrt{N}\,\varepsilon_n}{\sqrt{V}}\right)^V e^{-2(\sqrt{N}\,\varepsilon_n)^2}$$

where D is a constant that only depends on C. The remainder of the proof is exactly as in the proof of Theorem 5. Using (2.16), one gets finally the exponential bound

$$P(\|\pi_n^N-\pi_n\|_{\mathscr{F}}>\varepsilon)\leqslant(n+1)\left(\frac{D\sqrt{N}\,\varepsilon_n}{\sqrt{V}}\right)^V e^{-2(\sqrt{N}\,\varepsilon_n)^2}$$

Hence, if we denote $\delta=\sqrt{N}\,\varepsilon/\sigma_n$ we obtain the desired inequality and the theorem is thus established. □

3.2. A Uniform Convergence Result

Here we discuss the long time behavior of the particle scheme (1). In a previous work,[16] it was shown that under some conditions on the mappings $\{\phi_n; n \geqslant 1\}$, there exists a convergence rate $\alpha > 0$ such that for any bounded test function f

$$\sup_{n \geqslant 0} E(|\pi_n^N(f) - \pi_n(f)|) \leqslant \frac{C}{N^{\alpha/2}} \|f\| \tag{2.17}$$

for some constant C that does not depend on the time parameter $n \geqslant 0$.

The aim of this section is to turn this result into a statement uniform in f varying in a suitable class of function $\mathscr{F}$.

Before starting this discussion, let us present some comments on the results of Ref. 16. The main idea in this work was to connect the asymptotic stability of the limiting system (1.2) with the study of the long time behavior of the corresponding particle scheme. One condition under which the system (1.2) is asymptotically stable is the following

(H5) There exists a reference probability measure $\lambda \in \mathbf{M}_1(E)$ and $\eta \in (0, 1]$ such that $K_n(x, \cdot) \sim \lambda$ for any $x \in E$ and $n \geqslant 1$ and

$$\eta \leqslant \frac{dK_n(x, \cdot)}{d\lambda} \leqslant \frac{1}{\eta}$$

More precisely, under (H5) one can show that for any $n \geqslant 0$

$$\|\phi_{n|0}(\mu) - \phi_{n|0}(\mu)\|_{TV} \leqslant (1-\eta^2)^n, \qquad \forall \mu, \nu \in \mathbf{M}_1(E) \tag{2.18}$$

where $\|\cdot\|_{TV}$ denotes the total variation norm on $\mathbf{M}_1(E)$.

It should be mentioned that (H5) is stronger than the condition (H3) needed in the identification of the limiting Gaussian process in Donsker's theorem. Moreover, under appropriate regularity conditions on the fitness functions $\{g_n; n \geqslant 1\}$, the uniform estimate (2.17) holds for some convergence exponent $\alpha > 0$ which depends on the parameter η in (H5). Several ways to relax (H5) are also presented [see Del Moral and Guionnet[16]].

With these preliminaries out of the way, we now state and prove the following theorem.

Theorem 8. Let $\mathscr{F}$ be a countable collection of functions f such that $\|f\| \leqslant 1$ and satisfying (H5). Assume moreover that the limiting dynamical system (1.2) is asymptotically stable in the sense that, for some $\gamma > 0$,

$$\sup_{p \geqslant 0} |\phi_{p+T|p}(\mu)(f) - \phi_{p+T|p}(\nu)(f)| \leqslant e^{-\gamma T} \tag{2.19}$$

for all $\mu, \nu \in \mathbf{M}_1(E)$, $T \geqslant 0$ and $f \in \mathscr{F}$. When the fitness functions $\{g_n; n \geqslant 1\}$ satisfy (1.3) with $\sup_{n \geqslant 1} a_n = a < \infty$, then we have the following uniform estimate with respect to time

$$\sup_{n \geqslant 0} E(\|\pi_n^N - \pi_n\|_{\mathscr{F}}) \leqslant \frac{Ce^{\gamma'}}{N^{\alpha/2}} I(\mathscr{F}) \tag{2.20}$$

where C is a universal constant and α and γ' are given by

$$\alpha = \frac{\gamma}{\gamma + \gamma'} \qquad \text{and} \qquad \gamma' = 1 + 2 \log a$$

Proof. We use again the decomposition (2.4). By the same line of arguments as the ones given in the proof of Lemma 5, and recalling that π_p^N is the empirical measure associated to N conditionally independent random variables with common law $\phi_p(\pi_{p-1}^N)$ one can prove the error bound

$$E(\|\phi_{n|p}(\pi_p^N) - \phi_{n|p}(\phi_p(\pi_{p-1}^N))\|_{\mathscr{F}} \mid \pi_{p-1}^N) \leqslant \frac{C}{\sqrt{N}} a_{n|p}^2 I(\mathscr{F}_{n|p})$$

where $C > 0$ is a universal constant. Since $a_{n|p} \leqslant a_{n|0} \leqslant a^n$, $0 \leqslant p \leqslant n$, by Lemma 3, $I(\mathscr{F}_{n|p}) \leqslant I(\mathscr{F})$, and one concludes that

$$E(\|\phi_{n|p}(\pi_p^N) - \phi_{n|p}(\phi_p(\pi_{p-1}^N))\|_{\mathscr{F}}) \leqslant \frac{C}{\sqrt{N}} a^{2n} I(\mathscr{F})$$

Therefore, for any time $T \geqslant 0$, any $N \geqslant 1$ and any $f \in \mathscr{F}$,

$$\sup_{0 \leqslant n \leqslant T} E(\|\pi_n^N - \pi_n\|_{\mathscr{F}}) \leqslant \frac{C}{\sqrt{N}} (T+1)\, a^{2T} I(\mathscr{F}) \tag{2.21}$$

Similarly, for any $n \geqslant T$, we have

$$\begin{aligned}|\pi_n^N(f) - \pi_n(f)| \leqslant \sum_{p=n-T+1}^{n} & |\phi_{n|p}(\pi_p^N)(f) - \phi_{n|p}(\phi_p(\pi_{p-1}^N))(f)| \\ & + |\phi_{n/n-T}(\pi_{n-T}^N)(f) - \phi_{n/n-T}(\pi_{n-T})(f)|\end{aligned}$$

Under our assumptions, this implies that, for every $n \geqslant T$,

$$|\pi_n^N(f) - \pi_n(f)| \leqslant e^{-\gamma T} + \sum_{p=n-T+1}^{n} |\phi_{n|p}(\pi_p^N)(f) - \phi_{n|p}(\phi_p(\pi_{p-1}^N))(f)|$$

and using the same arguments as before one can prove that

$$\sup_{n \geqslant T} E(\|\pi_n^N - \pi_n\|_{\mathscr{F}}) \leqslant e^{-\gamma T} + \frac{C}{\sqrt{N}}(T+1)\, a^{2T} I(\mathscr{F}) \tag{2.22}$$

Combining the right-hand sides of (2.21) and (2.22) leads to a uniform L_1-error bound with respect to time in the form of the inequality

$$\sup_{n \geqslant 0} E(\|\pi_n^N - \pi_n\|_{\mathscr{F}}) \leqslant e^{-\gamma T} + \frac{C'}{\sqrt{N}} e^{\gamma' T} I(\mathscr{F})$$

where $\gamma' = 1 + 2\log a$ and $C' > 0$ is a universal constant. Obviously, if we choose

$$T = T(N) \stackrel{\text{def}}{=} \left[\frac{1}{2}\frac{\log N}{\gamma + \gamma'}\right] + 1$$

where $[r]$ denotes the integer part of $r \in \mathbb{R}$, we get that

$$\sup_{n \geqslant 0} E(\|\pi_n^N - \pi_n\|_{\mathscr{F}}) \leqslant \frac{1}{N^{\alpha/2}}(1 + e^{\gamma'} C' I(\mathscr{F}))$$

where $\alpha = \gamma/(\gamma + \gamma')$. This ends the proof of the theorem. □

4. APPLICATIONS

The interacting particle systems models presented earlier have many applications in biology, genetic algorithms, nonlinear filtering and bootstrap. In this short subsection, we briefly comment these examples of applications. The reader who wishes to know more about specific applications of these particle algorithms is invited to consult the referenced papers.

Weighted bootstrap methods are based on re-sampling from given weighted random measure. Roughly speaking, the key idea of bootstrap methods is the following: if the weighted random measure is close to a given nonrandom distribution, then the empirical measure associated to this re-sampling scheme should imitate the role of the weighted one [the book[1] includes a useful survey on this subject]. In our framework, we form at each time n, a weighted empirical measure

$$\psi_n(\pi_n^N) = \sum_{i=1}^{N} \frac{g_n(\xi_n^i)}{\sum_{j=1}^{N} g_n(\xi_n^j)} \delta_{\xi_n^i}$$

and the selection mechanism consists of re-sampling N independent particles $\{\hat{\xi}_n^i; 1 \leqslant i \leqslant N\}$ with this law so that to obtain a new probability measure with atoms of size integer multiples of $1/N$. In contrast to the classical bootstrap theory, one tries to approximate a measure valued dynamical system (1.2) and the weights are dictated by the dynamics structure of the former dynamical system.

The basic model for the general nonlinear filtering problem consists of a time inhomogeneous Markov process $\{X_n; n \geqslant 0\}$ taking values in a Polish space $(E, \mathbf{B}(E))$ and, a nonlinear observation process $\{Y_n; n \geqslant 0\}$ taking values in $\mathbb{R}^d$ for some $d \geqslant 1$. The classical filtering problem can be summarized as to find the conditional distributions

$$\pi_n(f) \stackrel{\text{def}}{=} E(f(X_n) \mid Y_1,..., Y_n), \qquad f \in \mathscr{C}_b(E), \quad n \geqslant 0 \tag{4.1}$$

It was proven in a rather general setting by Kunita[25] and Stettner[35] that, given a series of observations $Y = y$, the distributions $\{\pi_n; n \geqslant 0\}$ are solution of a discrete time measure valued dynamical system of the form (1.2). In this framework, the Markov kernels $\{K_n; n \geqslant 1\}$ correspond to the transition probability kernels of the signal process $\{X_n; n \geqslant 0\}$ and the fitness functions $\{g_n; n \geqslant 1\}$ are related to the observation noise source and also depends on the observation data. In such a context, the resulting interacting particle scheme can be viewed as a stochastic adaptative grid approximation of the non-linear filtering equation [see for instance Del Moral[11–13] and references therein]. Because of its importance in practical situations, we devote the last subsection of this paper to this theme.

In biology, the former model is used to mimic the genetic process of biological organisms and more generally natural evolution processes. Most of the terminology we have used is drawn from this framework. For instance, in gene analysis, each population represents a chromosome and each individual particle is called a gene. In this setting the fitness function is usually time-homogeneous and it represents the performance of the set of genes in a chromosome.[22] These particle algorithms are also used in population analysis to model changes in the structure of population in time and in space.

The interacting particle system model (1.1) is not only designed to solve the nonlinear filtering equation or to mimic natural evolution. Several practical problems have been solved using this approach, including numerical function optimization, image processing, combinatorial optimization tasks and machine learning.

Among the huge literature on genetic algorithm we refer to the papers.[3, 18–21, 28, 33, 39–43]

4.1. The Non-linear Filtering Problem

4.1.1. Introduction

The nonlinear filtering problem consists in recursively computing the conditional distributions of a nonlinear signal given its noisy observations. This problem has been extensively studied in the literature and, with the notable exception of the linear-Gaussian situation or wider classes of models (Bènes filters[2]), optimal filters do not have finitely recursive solutions (Chaleyat-Maurel and Michel[7]). Although Kalman filtering[23, 27] is a popular tool in handling estimation problems, its optimality heavily depends on linearity. When used for nonlinear filtering (Extended Kalman Filter), its performance relies on and is limited by the linearization performed on the model under investigation.

It has been recently emphasized that a more efficient way to solve numerically the filtering problem is to use random particle systems. That particle algorithms are gaining popularity is attested by the list of referenced papers. Instead of hand-crafting algorithms often on the basis of *adhoc* criteria, particle systems approaches provide powerful tools for solving a large class of nonlinear filtering problems. Several practical problems which have been solved using these methods are given in by Carvalho,[5] and Carvalho and Del Moral;[6] including Radar/Sonar signal processing and GPS/INS integrations. Other comparisons and examples where the extended Kalman filter fails can be found in Bucy.[4] The present paper is concerned with the genetic type interacting particle systems introduced by Del Moral.[12] Not all the nonlinear filtering problems which arise in applications can be represented by (1.2). Here we assumed that all processes are time continuous and the observation noise source isle independent of the signal process. A new interacting particle scheme for solving continuous time filtering problems in which the noise source is correlated to the signal is proposed by Del Moral *et al.*[17] Several variants of the genetic-type interacting particle scheme studied in this paper have been presented by Crisan and Lyons;[8] and Crisan *et al.*[9, 10] These variants are less "time consuming" but as a result the size of the system is not fixed but random. It is shown[16] that one cannot expect a uniform convergence result with respect to time. Roughly speaking, the size of the system behaves as a martingale and, trivial cases apart, its increasing predictable quadratic variation is usually not uniformly integrable with respect to time. On the other hand, the central limit theorem for such variants with random population size is still an open subject of investigation.

4.1.2. Description of the Models

Let (X, Y) be a time-inhomogeneous and discrete time Markov process taking values in a product space $E \times \mathbb{R}^d$, $d \geqslant 1$, and defined by the system

$$\begin{cases} X = (X_n)_{n \geqslant 0} \\ Y_n = h_n(X_{n-1}) + V_n, \qquad n \geqslant 1 \end{cases} \tag{4.2}$$

where E is a Polish space, $h_n\colon E \to \mathbb{R}^d$, $d \geqslant 1$, are bounded continuous functions and V_n are independent random variables with continuous and positive density g_n with respect to Lebesgue measure. The signal process X that we consider is assumed to be a noninhomogeneous and E-valued Markov process with Feller transition probability kernel K_n, $n \geqslant 1$, and initial probability measure π_0, on E. We will also assume the observation noise V and X are independent.

The classical filtering problem is concerned with estimating the distribution of X_n conditionally to the observations up to time n. Namely,

$$\pi_n(f) \overset{\text{def}}{=} E(f(X_n) \mid Y_1, \ldots, Y_n)$$

for all $f \in \mathscr{C}_b(E)$. For a detailed discussion of the filtering problem, the reader is referred to the pioneering paper[36] and to the more rigorous studies.[24, 34] Recent developments can be found in Ocone[29] and Pardoux.[30]

The dynamical structure of the conditional distribution $\{\pi_n; n \geqslant 0\}$ is given by the following lemma.

Lemma 8. Given a fixed observation record $Y = y$, $\{\pi_n; n \geqslant 0\}$ is solution of the $\mathbf{M}_1(E)$-valued dynamical system

$$\pi_n = \phi_n(y_n, \pi_{n-1}), \qquad n \geqslant 1 \tag{4.3}$$

where $y_n \in \mathbb{R}^d$ is the current observation and ϕ_n is the continuous function given by

$$\phi_n(y_n, \eta) = \psi_n(y_n, \eta)\, K_n$$

where for any $f \in \mathscr{C}_b(E)$, $\pi \in \mathbf{M}_1(E)$ and $y \in \mathbb{R}^d$,

$$\psi_n(y, \pi)\, f = \frac{\int f(x)\, g_n(y - h_n(x))\, \pi(dx)}{\int g_n(y - h_n(z))\, \pi(dz)}$$

In view of (1.1), the transition probability kernels of the genetic scheme associated (4.3) are now given by

$$P_y(\xi_n \in dx \mid \xi_{n-1} = z) = \prod_{p=1}^{N} \sum_{i=1}^{N} \frac{g_n(y_n - h_n(z^i))}{\sum_{j=1}^{N} g_n(y_n - h_n(z^j))}\, K_n(z^i, dx^p) \tag{4.4}$$

Thus, we see that the particles move according the following rules

1. **Updating**: When the observation $Y_n = y_n$ is received, each particle examines the system of particles $\xi_{n-1} = (\xi_{n-1}^1,..., \xi_{n-1}^N)$ and chooses randomly a site ξ_{n-1}^i with probability

$$\frac{g_n(y_n - h_n(\xi_{n-1}^i))}{\sum_{j=1}^{N} g_n(y_n - h_n(\xi_{n-1}^j))}$$

2. **Prediction**: After the updating mechanism, each particle evolves according to the transition probability kernel of the signal process.

4.1.3. Applications

4.1.3.1. Quenched Results. In order to apply the results of the preceding sections to this setting, rather than (1.3) we simply use the following assumption

(F) For any time $n \geqslant 0$, h_n is bounded and continuous and g_n is continuous with strictly positive values.

We note that if (F) holds then it is easily seen that there exists a family of positive functions $\{a_n;\, n \geqslant 0\}$, such that, for all $(y, x) \in \mathbb{R}^d \times E$ and $n \geqslant 0$,

$$a_n(y)^{-1} \leqslant \frac{g_n(y - h_n(x))}{g_n(y_n)} \leqslant a_n(y)$$

In such a framework, the fitness functions depend on the observation record. The assumption (F) is chosen so that the bounds (1.3) hold for some constants depending on the observation parameters. We also note that we can replace the fitness functions $g_n(y - h_n(\cdot))$ by the "normalized" ones $g_n(y)^{-1}\, g_n(y - h_n(\cdot))$ without altering the dynamical structure of Ocone[29] or Pardoux.[30] For any observation record $Y = y$, the results presented in Sections 2 and 3 hold although the constants $\{a_n; n \geqslant 1\}$ and the covariance function in Donsker's theorem will depend here on the observation parameters $\{y_n; n \geqslant 1\}$.

4.1.3.2. Averaged Results. One question that one may naturally ask is whether the averaged version of Theorems 2 and 3 hold. In many practical situations, the functions $a_n: \mathbb{R}^d \to \mathbb{R}_+$, $n \geqslant 1$, have a rather complicated form and it is difficult to obtain an averaged version of Theorem 3. Nevertheless, the averaged version of the L_1-error bounds presented in Theorem 4 does hold for a large class of non-linear sensors. Instead of (F), we will use the following stronger condition

(F)′ For any time $n \geqslant 1$, there exists a positive function $a_n: \mathbb{R}^d \to [1, \infty)$ and a non-decreasing function $\theta: \mathbb{R} \to \mathbb{R}$ such that for all $x \in E$ and $y \in \mathbb{R}^d$,

$$\frac{1}{a_n(y)} \leqslant \frac{g_n(y - h_n(x))}{g_n(y)} \leqslant a_n(y) \tag{4.5}$$

and

$$|\log a_n(y+u) - \log a_n(y)| \leqslant \theta(\|u\|)$$

The main simplification due to (F)′ is that now

$$a_n^2(Y_n) \leqslant a_n^2(V_n) \exp(2\theta(\|h_n\|))$$

In view of the proof of Theorem 4, one can check that (2.20) remains valid if we replace the condition $\sup_{n \geqslant 1} a_n < \infty$ by

$$L \overset{\text{def}}{=} \sup_{n \geqslant 1} \log E(a^2(V_n))^{1/2} < \infty \qquad \text{and} \qquad M \overset{\text{def}}{=} \sup_{n \geqslant 1} \|h_n\| < \infty$$

In this case (2.20) holds with $\gamma' = 1 + 2(L + \theta(M))$.

4.1.4. Examples

Let us now illustrate examples of non-linear observation noise sources for which (F)′ holds.

1. As a typical example of nonlinear filtering problem, assume the functions $h_n: E \to \mathbb{R}^d$, $n \geqslant 1$, are bounded and continuous and the densities g_n given by

$$g_n(v) = \frac{1}{((2\pi)^d |R_n|)^{1/2}} \exp\left(-\frac{1}{2} v' R_n^{-1} v\right)$$

where R_n is a $d \times d$ symmetric positive matrix. This correspond to the situation where the observations are given by

$$Y_n = h_n(X_{n-1}) + V_n, \qquad n \geqslant 1$$

where $(V_n)_{n \geqslant 1}$ is a sequence of $\mathbb{R}^d$-valued and independent random variables with Gaussian densities. After some easy manipulations one concludes that (F)′ hold with with

$$\log a_n(y) = \tfrac{1}{2} \|R_n^{-1}\| \; \|h_n\|^2 + \|R_n^{-1}\| \; \|h_n\| \; |y|$$

where $\|R_n^{-1}\|$ is the spectral radius of R_n^{-1}. In addition we have that

$$|\log a_n(y+u) - \log a_n(y)| \leqslant L_n \, |u| \qquad \text{with} \quad L_n = \|R_n^{-1}\| \; \|h_n\|$$

2. Our result is not restricted to Gaussian noise sources. For instance, let us assume that $d = 1$ and an is a bilateral exponential density

$$g_n(v) = \frac{\alpha_n}{2} \exp(-\alpha_n \, |v|), \qquad \alpha_n > 0$$

In this case (F)′ holds with

$$\log a_n(y) = \alpha_n \, \|h_n\|$$

which is independent of the observation parameter y. Another interesting remark is that in this particular case the exponential bounds of Theorem 3 hold.

REFERENCES

1. Barbe, P., and Bertail, P. (1995). The weighted bootstrap, *Lecture Notes in Statistics*, No. 98, Springer-Verlag.
2. Benes, B. E. (1981). Exact finite-dimensional filters for certain diffusions with nonlinear drift. *Stochastics* **5**, 65–92.
3. Bridges, C. L., and Goldberg, D. E. (1987). An analysis of reproduction and crossover in a binary-coded genetic algorithm. In Grefenstette, J. J. (ed.), *Proc. Second Int'l. Conf. Genetic Algorithms*, Lawrence Erlbaum Associates.
4. Bucy, R. S. (1994). *Lectures on Discrete Time Filtering*, Springer Verlag, Signal Processing and Digital Filtering.
5. Carvalho, H. (1995). Filtrage optimal nonlinéaire du signal GPS NAVSTAR en racalage de centrales de navigation, Thèse de L'Ecole Nationale Supérieure de l'Aéronautique et de l'Espace, Toulouse.
6. Carvalho, H., Del Moral, P., Monin, A., and Salut, G. (1997). Optimal nonlinear filtering in GPS/INS integration. *IEEE Trans. Aerospace Electronic Syst.* **33**(3), 835–850.
7. Chaleyat-Maurel, M., and Michel, D. (1983). Des résultats de nonexistence de filtres de dimension finie, *C. R. Acad. Sci. Paris*, Serie I, t. 296.

8. Crisan, D., and Lyons, T. J. (1997). Nonlinear filtering and measure valued processes. *Proc. Th. Rel. Fields* **109**, 217–244.
9. Crisan, D., Gaines, J., and Lyons, T. J. (1998). Convergence of a branching particle method to the solution of the zakai equation. *Society for Industrial and Applied Mathematics. J. Appl. Math.* **58**(5), 1568–1590.
10. Crisan, D., Del Moral, P., and Lyons, T. J. (1999). Nonlinear filtering using branching and interacting particle systems. *Markov Processes and Related Fields* **5**(3), 293–319.
11. Del Moral, P. (1998). Measure valued processes and interacting particle systems. Application to nonlinear filtering problems. *Ann. Appl. Prob.* **8**, 438–495.
12. Del Moral, P. (1996). Nonlinear filtering: Interacting particle resolution. *Markov Proc. Rel. Fields* **2**, 555–581.
13. Del Moral, P. (1997). Filtrage nonlinéaire par systèmes de particules en intéraction, *C. R. Acad. Sci. Paris*, t. 325, Série I, pp. 653–658.
14. Del Moral, P., and Guionnet, A. (1999). A central limit theorem for nonlinear filtering using interacting particle systems. *Ann. Appl. Prob.* (to appear).
15. Del Moral, P., and Guionnet, A. (1998). Large deviations for interacting particle systems. Applications to nonlinear filtering problems. *Stoch. Proc. Appl.* **78**, 69–95.
16. Del Moral, P., and Guionnet, A. (1998). On the stability of measure valued processes. Applications to nonlinear filtering and interacting particle systems, Publications du Laboratoire de Statistiques et Probabilités, Université Paul Sabatier, No. 03–98.
17. Del Moral, P., Jacod, J., and Protter, Ph. (1998). The Monte-Carlo method for filtering with discrete time observations. Publications du Laboratoire de Probabilités, Paris VI.
18. Goldberg, D. E. (1985). Genetic algorithms and rule learning in dynamic control systems, *Proc. First Int'l. Conf. Genetic Algorithms*, Lawrence Erlbaum Associates Hillsdale, New Jersey, pp. 8–15.
19. Goldberg, D. E. (1987). Simple genetic algorithms and the minimal deceptive problem. In Lawrence Davis (ed.), *Genetic Algorithms and Simulated Annealing*, Pitman.
20. Goldberg, D. E. (1989). *Genetic Algorithms in Search, Optimization and Machine Learning*, Addison-Wesley, Reading, Massachusetts.
21. Gordon, N. J., Salmon, D. J., and Smith, A. F. M. (1993). Novel approach to nonlinear/non-Gaussian Bayesian state estimation, *IEEE*.
22. Holland, J. H. (1975). *Adaptation in Natural and Artificial Systems*, University of Michigan Press, Ann Arbor.
23. Jazwinski, A. H. (1970). *Stochastic Processes and Filtering Theory*, Academic Press.
24. Kallianpur, G., and Striebel, C. (1967). Stochastic differential equations occuring in the estimation of continuous parameter stochastic processes, Technical Report No. 103, Department of statistics, University of Minnesota.
25. Kunita, H. (1971). Asymptotic behavior of nonlinear filtering errors of Markov processes. *J. Multivar. Anal.* **1**(4), 365–393.
26. Ledoux, M., and Talagrand, M. (1991). *Probability in Banach Spaces*, Springer-Verlag.
27. Liang, D., and McMillan, J. (1989). Development of a marine integrand navigation system, *Kalman Filter Integration of Modern Guidance and Navigation Systems*, AGARD-LS-166, OTAN.
28. Nix, A., and Vose, M. D. (1991). Modeling genetic algorithms with Markov chains. *Ann. Math. Artificial Intelligence* **5**, 79–88.
29. Ocone, D. L. (1980). Topics in nonlinear filtering theory, Ph.D. thesis, MIT, Cambridge.
30. Pardoux, E. (1989). Filtrage Nonlinéaire et Equations aux Dérivés Partielles Stochastiques Associées, Ecole d'été de Probabilités de Saint-Flour XIX. [*Lecture Notes in Mathematics*, 1464, Springer-Verlag. 1991.]

31. Pollard, D. (1984). *Convergence of Stochastic Processes*, Springer Verlag, New York.
32. Rachev, S. T. (1991). *Probability Metrics and the Stability of Stochastic Models*, Wiley, New York.
33. Segrest, P., and Goldberg, D. E. (1987). Finite Markov chain analysis of genetic algorithms. In Grefenstette, J. J. (ed.), *Proc. Second Int'l. Conf. Genetic algorithms*, Lawrence Erlbaum Associates.
34. Shiryaev, A. N. (1966). On stochastic equations in the theory of conditional Markov processes. *Theor. Prob. Appl.* **11**, 179–184.
35. Stettner, L. (1989). On Invariant Measures of Filtering Processes, *Stochastic Differential Systems*, *Lecture Notes in Control and Information Sciences*, Vol. 126, Springer-Verlag.
36. Stratonovich, R. L. (1960). Conditional Markov processes. *Theor. Prob. Appl.* **5**, 156–178.
37. Talagrand, M. (1994). Sharper bounds for Gaussian and empirical processes. *Ann. Prob.*, **22**, 28–76.
38. van der Vaart, A. N., and Wellner, J. A. (1996). Weak convergence and empirical processes with applications to statistics, *Springer Series in Statistics*.
39. Vose, M. D. (1993). Modeling simple genetic algorithms, *Foundations of Genetic Algorithms*, Morgan Kaufmann.
40. Vose, M. D. (1995). Modeling simple genetic algorithms. *Elementary Computations* **3**(4), 453–472.
41. Vose, M. D., and Liepins, G. E. (1993). Punctuated equilibria in genetic search. *Complex Systems*, **5**, 31–44.
42. Vose, M. D., and Wright, A. H. (1995). Simple genetic algorithms with linear fitness. *Elementary Computations* **2**(4), 347–368.
43. Whitley, L. D. (1993). An executable model of a simple genetic algorithm, *Foundations of Genetic Algorithms*, Morgan Kaufmann.