

## PARTICLE METHODS: AN INTRODUCTION WITH APPLICATIONS

PIERE DEL MORAL<sup>1</sup> AND ARNAUD DOUCET<sup>2</sup>

**Abstract.** Interacting particle methods are increasingly used to sample from complex high-dimensional distributions. They have found a wide range of applications in applied probability, Bayesian statistics and information engineering. Understanding rigorously these new Monte Carlo simulation tools leads to fascinating mathematics related to Feynman-Kac path integral theory and their interacting particle interpretations. In these lecture notes, we provide a pedagogical introduction to the stochastic modeling and the theoretical analysis of these particle algorithms. We also illustrate these methods through several applications including random walk confinements, particle absorption models, nonlinear filtering, stochastic optimization, combinatorial counting and directed polymer models.

**Résumé.** Les méthodes particulières en interaction sont de plus en plus utilisées pour simuler des mesures de probabilités complexes dans des espaces de grandes dimensions. Leurs domaines d'applications sont diverses et variés en probabilités appliquées, en statistique bayésienne et dans les sciences de l'ingénieur. L'analyse rigoureuse de ces nouvelles techniques de simulation de type Monte Carlo conduit à des techniques mathématiques fascinantes liées à la théorie des intégrales de Feynman et leurs interprétations particulières. Nous présentons dans ces notes une introduction pédagogique à la modélisation stochastique et l'analyse théorique de ces algorithmes particuliers. Nous illustrons ces modèles avec différentes applications, telles le confinement de marches aléatoires, des modèles d'évolutions de particules dans des milieux absorbants, des modèles de filtrage non linéaire, des problèmes d'optimisation stochastique, des questions de comptage combinatoire et des modèles de polymères dirigés.

### INTRODUCTION

Interacting particle methods are a class of Monte Carlo methods to sample from complex high-dimensional probability distributions and to estimate their normalizing constants. This class of algorithms approximate the target probability distributions by a large cloud of random samples termed particles. Practically, the particles evolve randomly around the space independently and to each particle is associated a non negative potential function. Periodically we duplicate particles with high potentials at the expense of particles with low potentials which die. This intuitive genetic mutation-selection type mechanism has appeared in numerous applications ranging from nonlinear filtering [3, 9, 10, 19–21, 23, 28, 34], Bayesian statistics [7, 11, 22, 35], combinatorial counting [1], molecular and polymer simulation [29], rare events simulation [5, 6, 26], quantum Monte Carlo methods [2, 31, 37] and genetic algorithms [27, 32] among others.

From a mathematical point of view, these methods can be interpreted as stochastic numerical approximations of Feynman-Kac measures. Feynman-Kac measures represent the distribution of the paths of a reference Markov

---

<sup>1</sup> Centre INRIA Bordeaux et Sud-Ouest & Institut de Mathématiques de Bordeaux, Université de Bordeaux I, 351 cours de la Libération 33405 Talence cedex, France, Pierre.Del-Moral@inria.fr

<sup>2</sup> Department of Statistics, Oxford University, doucet@stats.ox.ac.uk.

process, weighted by a collection of potential functions. These functional models are natural mathematical extensions of the traditional change of probability measures, commonly used in importance sampling. The particle interpretation consists in evolving a population of particles mimicking natural evolution mechanisms. During the mutation stage, the particles evolve independently of one another, according to the same probability transitions as the ones of the reference Markov chain. During the selection stage, each particle evaluates the potential value of its location. The ones with small relative values are killed, while the ones with high relative values are multiplied. The corresponding genealogical tree occupation measure converges, as the population size tends to infinity, to the complete Feynman-Kac distribution on path space.

The two authors and their collaborators started working on the design and the mathematical analysis of these algorithms in the mid 90's (see for instance [9, 10, 19], and references therein). Over the past few years, the popularity of these computationally intensive methods has dramatically increased thanks to the availability of cheap powerful computers. In particular in signal processing and machine learning, these algorithms are now widely used to solve nonlinear filtering problems. In this context, they are known as particle filters. The mathematical analysis of these algorithms offers a rigorous and unifying framework to analyze the convergence of numerous heuristic-like algorithms currently used in physics, statistics and engineering. It applies to any problem which can be translated in terms of functional Feynman-Kac type measures.

In this set of lecture notes, we provide a pedagogical introduction to the stochastic modeling and the theoretical analysis of these interacting particle algorithms. In a first section, section 1, we present several application areas and provide a detailed description of the corresponding interacting particle algorithms. Section 2 provides a brief treatise on Feynman-Kac modeling techniques. The last section, section 3, gives an overview of some convergence results, including variance and  $\mathbb{L}_p$ -mean error estimates, fluctuations and concentration properties. We have tried to give a brief "exposé" of the modern mathematical theory that is useful for the analysis of the asymptotic behavior of Feynman-Kac and particle models. To simplify the presentation and to clarify the main ideas behind these stochastic models, we have chosen to restrict these lectures notes to finite or countable state space models, avoiding any measure theory irrelevancies. In this simplified framework, we develop a rigorous mathematical analysis only involving vector and matrix operations. We emphasize that all of these particle models and the associated convergence results can be extended to general state-space models, including path-space models and excursion spaces on measurable state spaces.

We undertook this project for two main reasons:

First, we felt that there was no accessible and pedagogical treatment on stochastic particle models and their application areas. One objective in writing these lecture notes is to throw some new light on some interesting links between physical, engineering, statistical and mathematical domains which appear disconnected at first glance.

Second, the mathematical concepts and models are now at a point where they provide a very natural and unifying mathematical basis for a large class of Monte Carlo algorithms. We hope that this unifying point of view will help to develop fruitfully this field further.

## 1. EXAMPLES

The list of applications discussed here is by no means exhaustive and it just reflects the scientific interests of the authors.

### 1.1. Random walks confined in a set

We consider a symmetric random walk  $X_n$  on the integers  $\mathbb{Z}$  starting at the origin  $X_0 = 0$ . More formally, we take independent random variables  $U_n$ , where  $\mathbb{P}(U_n = 1) = \mathbb{P}(U_n = -1) = 1/2$  and we set  $X_n = X_0 + \sum_{1 \leq p \leq n} U_p$ . We fix  $A = \{-a + 1, -a + 2, \dots, a - 1\}$ , with  $a \in \mathbb{N}$ . We want to compute the conditional target distributions

$$\text{Law}((X_0, \dots, X_n) \mid \forall 0 \leq p \leq n, X_p \in A) \quad (1.1)$$

as well as the quantities

$$\mathcal{Z}_n := \mathbb{P}(\forall 0 \leq p < n, X_p \in A)$$

A naive Monte Carlo rejection type method consists with sampling  $N$  independent copies  $(X_n^i)_{1 \leq i \leq N}$  of the chain  $X_n$ . Rejecting the samples that exit the set  $A$ , we will have the estimates

$$\mathcal{Z}_n^N = N_n^A / N \simeq_{N \uparrow \infty} \mathcal{Z}_n \quad \text{with} \quad N_n^A := \sum_{1 \leq i \leq N} 1_{\{\forall 0 \leq p < n, X_p^i \in A\}} = \sum_{1 \leq i \leq N} \prod_{0 \leq p < n} 1_A(X_p^i)$$

and

$$\frac{1}{N_n^A} \sum_{1 \leq i \leq N} 1_{\{\forall 0 \leq p < n, X_p^i \in A\}} \delta_{(X_0^i, \dots, X_n^i)} \simeq_{N \uparrow \infty} \text{Law}((X_0, \dots, X_n) \mid \forall 0 \leq p < n, X_p \in A)$$

Notice that the l.h.s. of the above display is well defined as soon as the number of samples  $N_n^A$  remaining in  $A$  during  $n$  time step is not null. The main drawback of this rejection type Monte Carlo method comes from the fact that  $N_n^A = 0$ , after some very sort time period, in the sense that

$$\mathbb{P}(N_n^A > 0) = \mathbb{P}(\exists 1 \leq i \leq N : \forall 0 \leq p < n, X_p^i \in A) = 1 - (1 - \mathcal{Z}_n)^N \simeq 1 - e^{-N\mathcal{Z}_n}$$

for small values of  $\mathcal{Z}_n$ . In addition, we have

$$\mathbb{E} \left( \left( \frac{\mathcal{Z}_n^N}{\mathcal{Z}_n} - 1 \right)^2 \right) = \frac{1}{N\mathcal{Z}_n} (1 - \mathcal{Z}_n)$$

These observations show that the number of samples  $N$  needs to be much larger than  $\mathcal{Z}_n$  to have a well behaved approximation. For small values of  $a$ , the probability  $\mathcal{Z}_n$  for a random walk to remain confined in  $A$  for a long time is exponentially small w.r.t. the time parameter. This shows that the naive rejection Monte Carlo method cannot be used to approximate these quantities.

One way to solve this problem is to introduce correlations between the samples. The following mean field interacting type particle algorithm can be interpreted as a rejection type Monte Carlo scheme incorporating interacting recycling mechanisms.

We start with  $N$  particles at the origin denoted by  $\xi_0^i = 0$ , with  $i = 1, \dots, N$ . Each of them evolves  $\xi_0^i \rightsquigarrow \xi_1^i$  according to one transition of the random walk; more formally, we sample  $N$  independent copies  $(U_1^i)_{1 \leq i \leq N}$  of the random variables  $U_1$ , and we set  $\xi_1^i = \xi_0^i + U_1^i$ . We denote

$$\eta_1^N(1_A) = \frac{1}{N} \sum_{1 \leq i \leq N} 1_A(\xi_1^i) = \frac{1}{N} \text{Card} \{1 \leq i \leq N : \xi_1^i \in A\}$$

the proportion of points  $\xi_1^i$  in the set  $A$ . We define from the sample population  $(\xi_1^i)_{1 \leq i \leq N}$  a new population of  $N$  individuals  $(\widehat{\xi}_1^i)_{1 \leq i \leq N}$  as follows. For each  $i = 1, \dots, N$ , we perform the following operation: If  $\xi_1^i \in A$ ,

we set  $\widehat{\xi}_1^i = \xi_1^i$ . If  $\xi_1^i \notin A$ , we pick randomly an individual  $\tilde{\xi}_1^i$  among those  $\xi_1^j$  in the set  $A$  and we set  $\widehat{\xi}_1^i = \tilde{\xi}_1^i$ . In other words, individuals within  $A$  do not move, while the individuals outside  $A$  are replaced by a randomly chosen individual among those in the set  $A$ . It may happen that all individuals  $\xi_1^i$  are outside of the set  $A$ . In this case, the algorithm stops and we set  $\tau^N = 1$  to report the time of this event. If the algorithm has not stopped, we have a new configuration  $(\widehat{\xi}_1^i)_{1 \leq i \leq N}$  of  $N$  individuals in the set  $A$ . We evolve  $\widehat{\xi}_1^i \rightsquigarrow \xi_2^i$  according to one transition of the random walk; that is we sample  $N$  independent copies  $(U_2^i)_{1 \leq i \leq N}$  of the random variables  $U_2$ , we set  $\xi_2^i = \widehat{\xi}_1^i + U_2^i$  and we define

$$\eta_2^N(1_A) = \frac{1}{N} \sum_{1 \leq i \leq N} 1_A(\xi_2^i) = \frac{1}{N} \text{Card} \{1 \leq i \leq N : \xi_2^i \in A\}.$$

As before, we define from the sample population  $(\xi_2^i)_{1 \leq i \leq N}$  a new population of  $N$  individuals  $(\widehat{\xi}_2^i)_{1 \leq i \leq N}$ : individuals within  $A$  do not move, while the individuals outside the desired set are replaced by a randomly chosen individual among those in the set  $A$ . If all individuals  $\xi_2^i$  fall are outside of the set  $A$ , we set  $\tau^N = 2$ . Iterating this stochastic process, for every time  $n$  ( $< \tau^N$ ), we define a sequence of genetic type populations

$$\xi_n := (\xi_n^i)_{1 \leq i \leq N} \in \mathbb{Z}^N \xrightarrow{\text{selection}} \widehat{\xi}_n := (\widehat{\xi}_n^i)_{1 \leq i \leq N} \in \mathbb{Z}^N \xrightarrow{\text{mutation}} \xi_{n+1} \in \mathbb{Z}^N \quad (1.2)$$

This stochastic algorithm can be interpreted as a genetic type model with mutation transitions given by the one of a symmetric random walk and an acceptance-rejection selection type transition associated with the potential indicator type function  $1_A$ . Several estimates can be extracted from this interacting sampling algorithm.

First, we mention that the stopping time  $\tau^N$  tends to infinity as the size of the population  $N \rightarrow \infty$ . More precisely, the probability that the algorithm stops at a given time  $n$  tends to zero exponentially fast, as  $N$  tends to infinity. More interestingly, the product of the proportions of surviving particles at each time step

$$\mathcal{Z}_n^N := \prod_{0 \leq p < n} \eta_p^N(1_A)$$

is asymptotically a consistent estimate of the quantity  $P_n(A)$  and it is unbiased; that is we have

$$\lim_{N \rightarrow \infty} \mathcal{Z}_n^N = \mathcal{Z}_n \quad \text{and} \quad \mathbb{E}(\mathcal{Z}_n^N) = \mathcal{Z}_n \quad (1.3)$$

The convergence on the l.h.s. is an almost sure asymptotic convergence. It can be made precise by non asymptotic estimates including non asymptotic variance estimates and more refined exponential type deviations. If we interpret the selection transition as a birth and death process, then the important notion of the ancestral line of a current individual arises. More precisely, when a particle  $\widehat{\xi}_{n-1}^i \rightarrow \xi_n^i$  evolves to a new location  $\xi_n^i$ , we can interpret  $\widehat{\xi}_{n-1}^i$  as the parent of  $\xi_n^i$ . Looking backwards in time and recalling that the particle  $\widehat{\xi}_{n-1}^i$  has selected a site  $\xi_{n-1}^j$  in the configuration at time  $(n-1)$ , we can interpret this site  $\xi_{n-1}^j$  as the parent of  $\widehat{\xi}_{n-1}^i$  and therefore as the ancestor  $\xi_{n-1,n}^i$  at level  $(n-1)$  of  $\xi_n^i$ . Running back in time we can construct the whole ancestral line

$$\xi_{0,n}^i \leftarrow \xi_{1,n}^i \leftarrow \dots \leftarrow \xi_{n-1,n}^i \leftarrow \xi_{n,n}^i = \xi_n^i \quad (1.4)$$

of each current individual. The occupation measures of the corresponding  $N$ -genealogical tree model converge as  $N \rightarrow \infty$  to the conditional distribution (1.1). In a sense to be given, for any function  $f$  on the set  $\mathbb{Z}^{n+1}$ , we have the convergence, as  $N \rightarrow \infty$ ,

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N f(\xi_{0,n}^i, \xi_{1,n}^i, \dots, \xi_{n,n}^i) 1_{\tau^N > n} = \mathbb{E}(f(X_0, \dots, X_n) \mid \forall 0 \leq p < n, X_p \in A) \quad (1.5)$$

This convergence result can be refined in various directions. For instance, we can prove that the ancestral lines are ‘‘almost’’ independent with a common distribution given by the limiting conditional distribution. This property is often called the propagation of chaos property in applied probability. It refers to the fact that the initial population consists of independent and identically distributed random variables and that this property ‘‘propagates’’ approximately despite the introduction of interactions. Many other results can be derived including the fluctuations and the exponential concentration of the occupation measures of the genealogical tree around the limiting conditional distribution.

Besides the fact that the particle model approximate the (rare event) probabilities (1.3) and the conditional distributions (1.5) in path spaces, it also contains some information about the top of the spectrum of the matrix

$Q$  defined below

$$\forall(x, y) \in \{-a, -a + 1, \dots, a - 1, a\} \quad Q(x, y) := G(x) M(x, y)$$

with

$$G(x) := 1_A(x) \quad \text{and} \quad M(x, y) = \frac{1}{2} 1_{x-1}(y) + \frac{1}{2} 1_{x+1}(y)$$

Indeed, if we let  $\lambda$  the top eigenvalue of  $Q$  and we denote by  $h$  the corresponding eigenvector s.t.  $\sum_x h(x) = 1$ , then we have

$$\lim_{N \rightarrow \infty} \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{0 \leq p \leq n} \log \eta_p^N(1_A) = \log \lambda$$

In addition, the value  $h(x)$  coincides with the long time proportion of visits of the algorithm to the state  $x$ . In other words,  $h(x)$  can be interpreted as the limiting distribution of the individuals within the set  $A$ ; that is

$$\lim_{N, n \rightarrow \infty} \frac{1}{n} \sum_{0 \leq p \leq n} \frac{1}{N} \sum_{1 \leq i \leq N} 1_x(\hat{\xi}_n^i) 1_{\tau^N > n} = h(x) = \lim_{N, n \rightarrow \infty} \frac{1}{N} \sum_{1 \leq i \leq N} 1_x(\hat{\xi}_n^i) 1_{\tau^N > n}$$

The particle approximation model discussed above is far from unique. Many other interacting sampling strategies can be introduced by a simple change of probability measure. For instance, we can replace the mutation or the free evolution of the individuals in the previous algorithm by local moves restricted to the desired set  $A$ . These mutation type transitions  $\hat{\xi}_{n-1} \rightsquigarrow \xi_n$  can also be seen as transitions of a simple random walk on  $\mathbb{Z}$  reflected at the boundaries of the set  $A$ . By construction all the individuals  $\xi_n^i$  at any time horizon  $n$  and for any index  $i = 1, \dots, N$  are in the desired set  $A$ .

The corresponding selection transition  $\xi_n \rightsquigarrow \hat{\xi}_n$  is now defined as follows: Each individual  $\xi_n^i = x$  on the boundary  $x \in \partial A = \{-a + 1, (a - 1)\}$  of the set  $A$  has a probability  $G(x) := 1/2$  to stay in  $A$ , while the other individuals  $\xi_n^i$  (which are in the set  $A$ ) have a probability  $G(x) = 1$  to stay in  $A$ . The population  $\hat{\xi}_n$  is now defined as follows. For every index  $i$ , with a probability  $G(\xi_n^i)$ , we set  $\hat{\xi}_n^i = \xi_n^i$ , otherwise we replace  $\xi_n^i$  be a new individual  $\hat{\xi}_n^i = \xi_n^j$  randomly chosen in the whole population with a probability proportional to  $G(\xi_n^j)$ . If we now write  $\eta_n^N(G) = \frac{1}{N} \sum_{1 \leq i \leq N} G(\xi_n^i)$ , all the previous particle approximation results (corresponding to  $G(x) = 1_A(x)$ ) we have presented remain valid for this new particle algorithm.

## 1.2. Particle absorption models

The sampling techniques described in section 1.1 are far from being restricted to random walks models confined to a set. These strategies apply to a variety of application areas including computational physics, nonlinear filtering, biology, as well as in rare event analysis. From the pure mathematical point of view, they correspond to interacting particle approximation models of Feynman-Kac measures in path spaces.

To introduce these models, we recall that the conditional distributions discussed in (1.1) can be represented in terms of the distributions of the free path evolutions

$$\begin{aligned} \mathbb{P}_n(x_0, \dots, x_n) &= \text{Proba}((X_0, \dots, X_n) = (x_0, \dots, x_n)) \\ &= 1_0(x_0) M_1(x_0, x_1) \dots M_n(x_{n-1}, x_n) \end{aligned} \tag{1.6}$$

of the simple random walk starting at the origin with elementary transitions given by the matrix  $M_n := (M_n(x, y))_{x, y \in \mathbb{Z}}$  with entries given by

$$M_n(x, y) := \frac{1}{2} 1_{x-1}(y) + \frac{1}{2} 1_{x+1}(y)$$

More formally, if we set

$$\mathbb{Q}_n(x_0, \dots, x_n) := \text{Proba}((X_0, \dots, X_n) = (x_0, \dots, x_n) \mid \forall 0 \leq p < n, X_p \in A)$$

then we have

$$\mathbb{Q}_n(x_0, \dots, x_n) = \frac{1}{\mathcal{Z}_n} \left\{ \prod_{0 \leq p < n} G_p(x_p) \right\} \mathbb{P}_n(x_0, \dots, x_n) \quad (1.7)$$

with the indicator potential functions  $G_n(x) = 1_A(x)$  and  $\mathbb{P}_n(x_0, \dots, x_n)$  being the distribution of a free path of length  $n$  of the symmetric random walk. In (1.7),  $\mathcal{Z}_n$  is the normalizing constant given by

$$\mathcal{Z}_n = \mathbb{P}(\forall 0 \leq p < n, X_p \in A) = \mathbb{E} \left( \prod_{0 \leq p < n} G_p(X_p) \right)$$

These path integration type models are called Feynman-Kac measures in reference to Feynman path integral formulation of quantum mechanics where the classical notion of a single deterministic trajectory for a system is replaced by a sum over all possible trajectories weighted by the contributions of all the histories in configuration space. The Feynman-Kac measures presented in (1.7) can be regarded as the distribution of the paths of a Markov particle evolving using the Markov transitions  $M_n$  in an environment with absorbing obstacles related to potential functions  $G_n$ , and starting with some initial distribution  $\text{Law}(X_0) = \eta_0$  with  $\eta_0(x_0) = 1_0(x_0)$  in (1.6). To be more precise, we consider an auxiliary coffin or cemetery state  $c$  and we set  $E_c = E \cup \{c\}$ . We define an  $E_c$ -valued Markov chain  $X_n^c$  with two separate killing/exploration transitions:

$$X_n^c \xrightarrow{\text{killing}} \widehat{X}_n^c \xrightarrow{\text{exploration}} X_{n+1}^c \quad (1.8)$$

This killing/exploration mechanism are defined as follows:

- **Killing:** If  $X_n^c = c$ , we set  $\widehat{X}_n^c = c$ . Otherwise the particle  $X_n^c$  is still alive. In this case, with a probability  $G_n(X_n^c)$ , it remains in the same site so that  $\widehat{X}_n^c = X_n^c$ , and with a probability  $1 - G_n(X_n^c)$  it is killed and we set  $\widehat{X}_n^c = c$ .
- **Exploration:** Once a particle has been killed, it cannot being brought back to life so if  $\widehat{X}_n^c = c$  then we set  $\widehat{X}_p^c = X_p = c$  for any  $p > n$ . Otherwise, the particle  $\widehat{X}_n^c \in E$  evolves to a new location  $X_{n+1}^c = x$  in  $E$  randomly chosen according to the distribution  $M_{n+1}(X_n^c, x)$ .

In this physical interpretation, the measure  $\mathbb{Q}_n$  represent the conditional distributions of the paths of a non absorbed Markov particle. To see this claim, we denote by  $T$  the time at which the particle has been killed

$$T = \inf \{n \geq 0 ; \widehat{X}_n^c = c\}$$

By construction, we have

$$\begin{aligned} & \text{Proba}(T > n - 1) \\ &= \text{Proba}(\widehat{X}_0^c \in E, \dots, \widehat{X}_{n-1}^c \in E) \\ &= \int_{E^{(n+1)}} \eta_0(dx_0) G_0(x_0) M_1(x_0, dx_1) \dots M_{n-1}(x_{n-2}, dx_{n-1}) G_{n-1}(x_{n-1}) \\ &= \mathbb{E} \left( \prod_{p=0}^{n-1} G_p(X_p) \right) \end{aligned}$$

This also shows that the normalizing constants  $\mathcal{Z}_n$  represent respectively the probability for the particle to be alive at time  $n - 1$ . In other words, we have that

$$\mathcal{Z}_n = \text{Proba}(T > n - 1)$$

Similar arguments yield that is the distribution of a particle conditional upon being alive at time  $n - 1$

$$\mathbb{Q}_n(x_0, \dots, x_n) = \text{Proba}((X_0^c, \dots, X_n^c) = (x_0, \dots, x_n) \mid T > n - 1)$$

The particle sampling technique of any distribution  $\mathbb{Q}_n$  associated with some Markov transition  $M_n$  and some sequence of  $[0, 1]$ -valued potential function  $G_n$  on some (countable) state space  $E$  is defined as before in terms of a genetic type algorithm with  $M_n$ -mutations and  $G_n$ -selection type transitions. More precisely, at every time step  $n$ , we sample the mutation-selection transitions as follows: During the mutation step, every individual performs a local random move according to the Markov transition  $M_n$ . During the selection step, every individual evaluates its potential value  $G_n(\xi_n^i)$ , with  $1 \leq i \leq N$ . For every index  $i$ , with a probability  $G_n(\xi_n^i)$ , we set  $\widehat{\xi}_n^i = \xi_n^i$ , otherwise we replace  $\xi_n^i$  by a fresh new individual  $\widehat{\xi}_n^i = \xi_n^j$  randomly chosen in the whole population with a probability proportional to  $G_n(\xi_n^j)$ . Notice that the number of individual remains constant.

More precisely, for any time horizon  $n$  and any function  $f$  on the set  $E^{n+1}$ , we have

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N f(\xi_{0,n}^i, \xi_{1,n}^i, \dots, \xi_{n,n}^i) 1_{\tau^N > n} = \sum_{x_0, \dots, x_n} f(x_0, \dots, x_n) \mathbb{Q}_n(x_0, \dots, x_n) \tag{1.9}$$

Furthermore, the  $N$ -particle approximation of the normalizing constants  $\mathcal{Z}_n$  are given by

$$\mathcal{Z}_n^N := \prod_{0 \leq p < n} \eta_p^N(G_p) \quad \text{with} \quad \forall n \in \mathbb{N} \quad \eta_n^N(G_n) := \frac{1}{N} \sum_{1 \leq i \leq N} G_n(\xi_n^i) \tag{1.10}$$

For time homogeneous models  $(G_n, M_n) = (G, M)$  associated with a reversible matrix  $M$  w.r.t. to some measure  $\lambda$  on  $E$ , i.e.  $\lambda(x) M(x, y) = \lambda(y) M(y, x)$ , the corresponding particle model also contains information about the top of the spectrum of the matrix  $Q$  defined through

$$\forall (x, y) \in E \quad Q(x, y) := G(x) M(x, y)$$

More precisely, if we let  $\lambda$  the top eigenvalue of  $Q$  in  $\mathbb{L}_2(\lambda)$  and we denote by  $h$  the corresponding eigenvector s.t.  $\sum_x \lambda(x) h(x) = 1$ , then we have

$$\lim_{N \rightarrow \infty} \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{0 \leq p \leq n} \log \eta_p^N(G) = \log \lambda$$

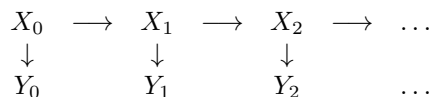
as well as

$$\lim_{N, n \rightarrow \infty} \frac{1}{n} \sum_{0 \leq p \leq n} \frac{1}{N} \sum_{1 \leq i \leq N} 1_x(\widehat{\xi}_n^i) 1_{\tau^N > n} = \lambda(x) h(x) = \lim_{N, n \rightarrow \infty} \frac{1}{N} \sum_{1 \leq i \leq N} 1_x(\widehat{\xi}_n^i) 1_{\tau^N > n}$$

For further details on this subject, we refer the reader to [9, 10, 16, 17] and references therein.

### 1.3. Nonlinear filtering problems

We discuss here the application of these particle model to filtering problems. Suppose that at every time step the state of the Markov chain  $X_n$  is partially observed according to the following schematic picture



with some random variables  $Y_n$  whose values only depend on the current state of the chain

$$\text{Proba}(Y_n = y_n \mid X_n = x_n) = G(x_n, y_n)$$

We consider the following pair of events

$$A_n(x) := \{(X_0, \dots, X_n) = (x_0, \dots, x_n)\} \quad \text{and} \quad B_{n-1}(y) := \{(Y_0, \dots, Y_{n-1}) = (y_0, \dots, y_{n-1})\}$$

The filtering problem consists of computing the conditional distributions of the state variables  $A_n(x)$  given the observations  $B_n(y)$ . By construction, given  $A_n(x)$ , the random variables are independent and identically distributed with a distribution given by

$$\text{Proba}(B_{n-1}(y) \mid A_n(x)) = \prod_{0 \leq p < n} G(x_p, y_p)$$

By direct application of Bayes' rule we have the following formula

$$\begin{aligned} \text{Proba}(A_n(x) \cap B_{n-1}(y)) &= \text{Proba}(B_{n-1}(y) \mid A_n(x)) \times \text{Proba}(A_n(x)) \\ &= \left\{ \prod_{0 \leq p < n} G(x_p, y_p) \right\} \mathbb{P}_n(x_0, \dots, x_n) \end{aligned} \quad (1.11)$$

from which we conclude that

$$\text{Proba}(A_n(x) \mid B_{n-1}(y)) = \frac{1}{\mathcal{Z}_n(y)} \left\{ \prod_{0 \leq p < n} G(x_p, y_p) \right\} \mathbb{P}_n(x_0, \dots, x_n)$$

with the normalizing constants

$$\mathcal{Z}_n(y) := \text{Proba}(B_{n-1}(y)) = \sum_{x_0, \dots, x_n} \left\{ \prod_{0 \leq p < n} G(x_p, y_p) \right\} \mathbb{P}_n(x_0, \dots, x_n)$$

These Feynman-Kac formulae express the conditional distributions of the path sequence  $(X_0, \dots, X_n)$  as the distribution of the signal paths  $(X_0, \dots, X_n) = (x_0, \dots, x_n)$  weighted by the product of the likelihood functions  $G(x_p, y_p)$  from the origin  $p = 0$  up to time  $p = n$ . If we fix the observation sequence  $Y_n = y_n$  and set  $G_n(x_n) := G(x_n, y_n)$ , these measures have exactly the same form as the one presented in (1.7). The corresponding particle approximations are often referred as particle filters in signal processing and statistics (see for instance [19], and references therein). These particle algorithms can also be used to approximate the log-likelihood functions using (1.10); that is the log-likelihood

$$L_n(y) := \log \mathcal{Z}_n(y)$$

is approximated using

$$L_n^N(y) := \log \mathcal{Z}_n^N(y) = \sum_{0 \leq p < n} \log \eta_p^N(G_p).$$

#### 1.4. Stochastic optimization algorithms

Suppose we want to compute the global minima of a given non negative cost function  $V$  on some finite state space  $E$  equipped with the counting measure  $\lambda(x) := \frac{1}{\text{Card}(E)}$ . From the probabilistic point of view, this



problem amounts of sampling random states according to the Boltzmann-Gibbs distributions associated with a large inverse temperature parameter  $\beta$  and given below

$$\mu_\beta(x) := \frac{1}{\mathcal{Z}_\beta} e^{-\beta V(x)} \lambda(x) \quad \text{with} \quad \mathcal{Z}_\beta := \sum_x e^{-\beta V(x)} \lambda(x) \quad (1.12)$$

There is no loss of generality to assume that  $\inf_x V(x) = 0$  and for any state  $x \notin V_0 := V^{-1}(\{0\})$ ,  $V(x) \geq \delta$  for some  $\delta > 0$ . It follows that we have

$$\text{Card}(V_0) \leq \mathcal{Z}_\beta \leq \text{Card}(V_0) + \text{Card}(V_0^c) e^{-\beta\delta} \xrightarrow{\beta \uparrow \infty} \text{Card}(V_0)$$

and therefore

$$\lim_{\beta \rightarrow \infty} \mu_\beta(x) = \mu_\infty(x) := 1_{V_0}(x) / \text{Card}(V_0)$$

This simple observation shows that sampling according to  $\mu_\beta$  is roughly equivalent to that of sampling randomly an unknown state variable with minimal cost. For very large state spaces, it is typically impossible to sample from  $\mu_\beta$  directly.

The celebrated simulated annealing algorithm to sample from  $\mu_\infty$  consists of sampling approximately from a sequence of distributions  $\mu_{\beta_n}$  where  $\beta_n$  is a non-decreasing sequence going to  $\infty$ . The rationale is that it is “easier” to sample from  $\mu_\beta$  when  $\beta$  is small; if  $\beta = 0$  then  $\mu_0$  is the uniform counting measure on  $E$  from which it is trivial to sample. For  $\beta_n > 0$ , we sample approximately from each intermediate distribution  $\mu_{\beta_n}$  using Markov chain Monte Carlo (MCMC) sampling techniques; that is we select a transition matrix  $M_{\beta_n} = (M_{\beta_n}(x, y))_{x, y \in E}$  with left eigenvector  $\mu_{\beta_n}$  associated with the eigenvalue 1, that is

$$\sum_x \mu_{\beta_n}(x) M_{\beta_n}(x, y) = \mu_{\beta_n}(y)$$

The probabilistic interpretation of the above equation is as follows: pick randomly a state  $x$  with distribution  $\mu_{\beta_n}(x)$  and take a random transition  $x \rightsquigarrow y$  from the distribution  $M_{\beta_n}(x, y)$ , then the probability of being at state  $y$  is again  $\mu_{\beta_n}(y)$ . The literature on MCMC methods discusses numerous choices of transitions  $M_{\beta_n}$  satisfying this property. The most famous is the Metropolis-Hastings transition associated to a symmetric transition matrix  $K(x, y) = K(y, x)$  and defined by

$$\begin{aligned} & M_{\beta_n}(x, y) \\ &= K(x, y) \min(1, e^{-\beta_n(V(y)-V(x))}) + (1 - \sum_z K(x, z) \min(1, e^{-\beta_n(V(z)-V(x))})) 1_x(y) \end{aligned}$$

Using the fundamental ergodic theorem for regular Markov chains, starting from any initial state  $x_0$ , the  $n$ -th step of a run of the Markov chain with transitions  $M_{\beta_n}$  has a probability very close to  $\mu_{\beta_n}(y)$  of being at the site  $y$ , for a large  $n$ . Practically, we select  $\beta_1$  and we run the chain starting at  $X_0 = x_0$  for a large enough number of runs  $n_1$  such that the law of the state  $X_{n_1}$  is close to  $\mu_{\beta_1}$

$$X_0 = x_0 \xrightarrow{M_{\beta_1}} X_1 \xrightarrow{M_{\beta_1}} \dots \xrightarrow{M_{\beta_1}} X_{n_1} \quad \text{with } n_1 \text{ large enough s.t. } \text{Law}(X_{n_1}) \simeq \mu_{\beta_1}$$

Notice that the choice of  $n_1$  depends on  $\beta_1$ : the larger  $\beta_1$  is, the “peakier”  $\mu_{\beta_1}$  is and the larger  $n_1$  is. When the chain is stabilized, we choose a  $\beta_2 > \beta_1$  and we run the chain starting at  $X_{n_1}$  for a new large enough number of time steps  $n_2$  such that the law of the state  $X_{n_1+n_2}$  is close to  $\mu_{\beta_2}$

$$X_{n_1} \xrightarrow{M_{\beta_2}} X_{n_1+1} \xrightarrow{M_{\beta_2}} \dots \xrightarrow{M_{\beta_2}} X_{n_1+n_2} \quad \text{with } n_2 \text{ large enough s.t. } \text{Law}(X_{n_1+n_2}) \simeq \mu_{\beta_2}$$

The theoretical “optimal” inverse temperature parameter ensuring convergence in some sense of the Markov chain to  $\mu_\infty$  is logarithmic [30]. This amounts to say that we change by one unit the parameter  $\beta$  on every time interval with exponential length. This is unrealistic from a practical point of view.

We present now an alternative particle strategy for sampling random states according to the sequence of measures  $\mu_{\beta_n}$  associated with a given non decreasing sequence of inverse temperature parameters  $\beta_n$ . We suppose that  $\beta_0 = 0$  so that  $\mu_{\beta_0}$  coincides with the uniform counting measure on the set  $E$ . We start with  $N$  independent individuals  $\xi_0 := (\xi_0^i)_{1 \leq i \leq N}$  randomly chosen in  $E$  according to  $\mu_{\beta_0}$ . We perform a selection transition  $\xi_0 \rightsquigarrow \widehat{\xi}_0 := \left( \widehat{\xi}_0^i \right)_{1 \leq i \leq N}$  using the potential functions  $G_0$  defined by

$$G_0(x) = \exp(-(\beta_1 - \beta_0)V(x))$$

In other words, every individual evaluates its potential value  $G_0(\xi_0^i)$ . For every index  $i$ , with a probability  $G_0(\xi_0^i)$ , we set  $\widehat{\xi}_0^i = \xi_0^i$ , otherwise we replace  $\xi_0^i$  by a new individual  $\widehat{\xi}_0^i = \xi_0^j$  randomly chosen in the whole population with a probability proportional to  $G_0(\xi_0^j)$ . During the mutation step  $\widehat{\xi}_0 \rightsquigarrow \xi_1 := (\xi_1^i)_{1 \leq i \leq N}$ , every selected individual  $\widehat{\xi}_0^i$  performs a local random move  $\widehat{\xi}_0^i \rightsquigarrow \xi_1^i$  (independently of one another) according to the Markov transition  $M_{\beta_1}$ . Then, we perform another selection transition  $\xi_1 \rightsquigarrow \widehat{\xi}_1 := \left( \widehat{\xi}_1^i \right)_{1 \leq i \leq N}$  using the fitness functions  $G_1$  defined below:

$$G_1(x) = \exp(-(\beta_2 - \beta_1)V(x))$$

After this selection stage we mutate each selected individual using the Markov transition  $M_{\beta_2}$ , and so on. Iterating these transitions, we define a simple genetic model with mutations transitions  $M_{\beta_n}$  and selection fitness functions  $G_n$ :

$$\xi_n := (\xi_n^i)_{1 \leq i \leq N} \in E^N \xrightarrow{\text{selection}} \widehat{\xi}_n := \left( \widehat{\xi}_n^i \right)_{1 \leq i \leq N} \in E^N \xrightarrow{\text{mutation}} \xi_{n+1} \in E^N \quad (1.13)$$

This algorithm was first proposed in [11]. A variety of convergence results can be established for this algorithm. For instance, for any function  $f$  on  $E$  and any time horizon, we have

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{1 \leq i \leq N} f(\xi_n^i) = \sum_x \mu_{\beta_n}(x) f(x)$$

In addition, if we set  $\eta_n^N(G_n) := \frac{1}{N} \sum_{1 \leq i \leq N} G_n(\xi_n^i)$ , the unbiased  $N$ -particle approximation  $\mathcal{Z}_{\beta_n}^N$  of the normalizing constants  $\mathcal{Z}_{\beta_n}$  is given by

$$\mathcal{Z}_{\beta_n}^N := \prod_{0 \leq p < n} \eta_p^N(G_p) \xrightarrow{N \rightarrow \infty} \mathcal{Z}_{\beta_n}$$

The particle model described above can be extended in various ways. For instance, an adaptive version based on tuning temperature schedules with the number of Markov Chain Monte Carlo iterations can be easily developed replacing  $M_{\beta_n}$  by some  $m_n$ -iterated transitions  $M_{\beta_n}^{m_n}$ . Another strategy is to choose at every time step the next inverse temperature  $\beta_{n+1} (\geq \beta_n)$  s.t.  $\eta_n^N(e^{-(\beta_{n+1} - \beta_n)V})$  is below some given threshold  $\epsilon_n$  [12, 25].

This class of interacting simulated annealing algorithm clearly differs from the Metropolis type low and higher temperature exchanges used in Parallel tempering (a.k.a. replica exchange MCMC sampling) [8, 24, 38].

## 1.5. Combinatorial counting and sampling

Suppose we want to compute the cardinality of a given subset  $A$  of some finite state space  $E$  equipped with the counting measure  $\lambda(x) := \frac{1}{\text{Card}(E)}$ . Once again, from a probabilistic point of view, this problem is

equivalent to computing the normalizing constant of the following Boltzmann-Gibbs distribution

$$\mu_A(x) := \frac{1}{\mathcal{Z}_A} 1_A(x) \lambda(x) \quad \text{with} \quad \mathcal{Z}_A := \sum_x 1_A(x) \lambda(x) \quad (1.14)$$

To sample from  $\mu_A$  and compute  $\mathcal{Z}_A$ , the idea consists of selecting a judicious sequence of decreasing subsets  $A_n$  in such a way that it is easy to sample states in  $A_n$  starting from the set  $A_{n-1}$ . We suppose that  $A_0 = E$  so that  $\mu_{A_0}$  coincide with the uniform counting measure on the set  $E$ . The algorithm is thus very similar to the one described above for optimization.

For any set  $A_n$ , we introduce an MCMC transition matrix  $M_{A_n} = (M_{A_n}(x, y))_{x, y \in E}$  with left eigenvector  $\mu_{A_n}$  associated with the eigenvalue 1, that is

$$\sum_x \mu_{A_n}(x) M_{A_n}(x, y) = \mu_{A_n}(y)$$

A simple Metropolis-Hasting type transition associated with a symmetric transition matrix  $K(x, y) = K(y, x)$  is given by

$$M_{A_n}(x, y) = K(x, y) 1_{A_n}(y) + \left(1 - \sum_z K(x, z) 1_{A_n}(z)\right) 1_x(y)$$

The  $N$ -particle stochastic algorithm is defined as follows. We start with  $N$  independent random individuals  $\xi_0 := (\xi_0^i)_{1 \leq i \leq N}$  randomly chosen in  $E$  with  $\mu_{A_0}$ . We perform a selection transition  $\xi_0 \rightsquigarrow \widehat{\xi}_0 := (\widehat{\xi}_0^i)_{1 \leq i \leq N}$  using the fitness functions  $G_0 = 1_{A_1}$ . In other words, every individual in the set  $A_1$  stays in the same place  $\widehat{\xi}_0^i = \xi_0^i$ , otherwise we replace  $\xi_0^i$  by a fresh new individual  $\widehat{\xi}_0^i = \xi_0^j$  randomly chosen among the individuals  $\xi_0^j \in A_1$ . When no individuals  $\xi_0^j$  are in the set  $A_1$ , the algorithm stops and we set  $\tau^N = 0$ . Assuming that  $\tau^N > 0$ , during the mutation step  $\widehat{\xi}_0 \rightsquigarrow \xi_1 := (\xi_1^i)_{1 \leq i \leq N}$ , every selected individual  $\widehat{\xi}_0^i$  performs a local random move  $\widehat{\xi}_0^i \rightsquigarrow \xi_1^i$  (independently of one another) in the set  $A_1$  according to the Markov transition  $M_{A_1}$ . Then, we perform another selection transition  $\xi_1 \rightsquigarrow \widehat{\xi}_1 := (\widehat{\xi}_1^i)_{1 \leq i \leq N}$  using the fitness functions  $G_1 = 1_{A_2}$ . When no individuals  $\xi_1^j$  are in the set  $A_2$ , the algorithm stops and we set  $\tau^N = 1$ . After this selection stage we mutate each selected individual using the Markov transition  $M_{A_2}$ , and so on. For any function  $f$  on  $E$  and any time horizon  $n$ , we have

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{1 \leq i \leq N} f(\xi_n^i) 1_{\tau^N > n} = \sum_x \mu_{A_n}(x) f(x)$$

In addition, if we set  $\eta_n^N(G_n) := \frac{1}{N} \sum_{1 \leq i \leq N} G_n(\xi_n^i)$ , the proportion of individuals in  $A_{n+1}$  after the  $n$ -th mutation, the unbiased  $N$ -particle approximation  $\mathcal{Z}_{A_n}^N$  of the normalizing constants  $\mathcal{Z}_{A_n}$  is given by

$$\mathcal{Z}_{A_n}^N := \prod_{0 \leq p < n} \eta_p^N(G_p) \xrightarrow{N \rightarrow \infty} \mathcal{Z}_{A_n} = \text{Card}(A_n) / \text{Card}(E)$$

Replacing  $(e^{-(\beta_{n+1} - \beta_n)V}, M_{\beta_n})$  by  $(1_{A_{n+1}}, M_{A_n})$  we observe that the particle model described above has the same form as the one discussed in section 1.4. Rephrasing the remarks given at the end of section 1.4, it can also be extended in various ways. For instance, an adaptive version based on tuning temperature schedules with the number of Markov Chain Monte Carlo iterations can be easily developed replacing  $M_{A_n}$  by some  $m_n$ -iterated transitions  $M_{A_n}^{m_n}$ . Another strategy is to choose at every time step the next subset  $A_{n+1}(\subset A_n)$  s.t.  $\eta_n^N(1_{A_{n+1}})$  is below some given threshold  $\epsilon_n$  [12, 25].

### 1.6. Genetic search algorithms

We consider an energy function or a cost criteria  $V : x \in E \mapsto (x)$  on some finite state space  $E$  where we assume  $\inf_x V(x) = 0$  without loss of generality. The objective is to find the global minima points  $x^* \in E$  s.t.  $V(x^*) = \inf_{x \in E} V(x)$ . Let  $V^*$  denote the set of these points. We described in Section 1.4 an interacting particle algorithm to solve this problem which relies on interacting simulated annealing type chains. We present here the more standard genetic algorithm with mutation and proportional selection.

To construct this algorithm, we introduce a collection of Markov transitions  $M_n(x, y)$  from  $E$  into itself. This collection of transition matrices represents the probability  $M_n(x, y)$  that a individual at site  $x$  evolves to a new state  $x$  during the  $n$ -th mutation transition.

The genetic algorithm with  $N$  individuals is defined as follows. We start with  $N$  independent random individuals  $\xi_0 := (\xi_0^i)_{1 \leq i \leq N}$  randomly chosen in  $E$  with some distribution, say  $\eta_0$ . We perform a proportional type selection transition  $\xi_0 \rightsquigarrow \widehat{\xi}_0 := (\widehat{\xi}_0^i)_{1 \leq i \leq N}$  using the potential functions  $G_0(\xi_0^i) = \exp(-\beta_0 V(\xi_0^i))$ , where  $\beta_0 \geq 0$  is an inverse temperature parameter. In other words, with probability  $G_0(\xi_0^i)$  every individual stays in the same place  $\widehat{\xi}_0^i = \xi_0^i$ ; otherwise, we replace  $\xi_0^i$  by a new individual  $\widehat{\xi}_0^i = \xi_0^j$  randomly chosen among the individuals  $\xi_0^j$  with a probability proportional to its weight  $G_0(\xi_0^j)$ . Formally, we set

$$\widehat{\xi}_0^i = \epsilon_0^i \xi_0^i + (1 - \epsilon_0^i) \widetilde{\xi}_0^i$$

where  $\epsilon_0^i$  stands for a sequence of independent  $\{0, 1\}$ -valued Bernoulli random variables with distributions

$$G_0(\xi_0^i) := \text{Proba}(\epsilon_0^i = 1 \mid \xi_0) = 1 - \text{Proba}(\epsilon_0^i = 0 \mid \xi_0)$$

and  $\widetilde{\xi}_0 := (\widetilde{\xi}_0^i)_{1 \leq i \leq N}$  are independent, identically distributed and  $\{\xi_0^j, 1 \leq j \leq N\}$ -valued random variables with common distributions given for any index  $1 \leq i \leq N$  by

$$\forall 1 \leq j \leq N \quad \text{Proba}(\widetilde{\xi}_0^i = \xi_0^j \mid \xi_0) = G_0(\xi_0^j) / \sum_{1 \leq j \leq N} G_0(\xi_0^j)$$

During the mutation step  $\widehat{\xi}_0 \rightsquigarrow \xi_1 := (\xi_1^i)_{1 \leq i \leq N}$ , every selected individual  $\widehat{\xi}_0^i$  performs a local random move  $\widehat{\xi}_0^i \rightsquigarrow \xi_1^i$  (independently of one another) according to the Markov transition  $M_1$ . Then, we perform another proportional type selection transition  $\xi_1 \rightsquigarrow \widehat{\xi}_1 := (\widehat{\xi}_1^i)_{1 \leq i \leq N}$  using the potential functions  $G_1(\xi_1^i) = \exp(-\beta_1 V(\xi_1^i))$ , where  $\beta_1 \geq 0$  is another inverse temperature parameter, and so on. We define in this way a sequence of genetic type populations  $\xi_n, \widehat{\xi}_n$ , as in (1.2) and the corresponding genealogical tree model (1.4) associated with the ancestral lines  $(\xi_{p,n}^i)_{0 \leq p \leq n}$  of every  $i$ -th individuals after the  $n$ -th mutation. In the same way, running back in time we have the whole ancestral line

$$\widehat{\xi}_{0,n}^i \longleftarrow \widehat{\xi}_{1,n}^i \longleftarrow \dots \longleftarrow \widehat{\xi}_{n-1,n}^i \longleftarrow \widehat{\xi}_{n,n}^i = \widehat{\xi}_n^i \quad (1.15)$$

of every  $i$ -th individual after the  $n$ -th selection.

For any function  $f$  on  $E^{(n+1)}$  and any time horizon  $n$ , we can prove that

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N f(\widehat{\xi}_{0,n}^i, \widehat{\xi}_{1,n}^i, \dots, \widehat{\xi}_{n,n}^i) = \frac{\mathbb{E} \left( f_n(X_0, \dots, X_n) \exp \left( - \sum_{0 \leq p \leq n} \beta_p V(X_p) \right) \right)}{\mathbb{E} \left( \exp \left( - \sum_{0 \leq p \leq n} \beta_p V(X_p) \right) \right)}$$

In other words, the proportion of paths  $(\widehat{\xi}_{0,n}^i, \widehat{\xi}_{1,n}^i, \dots, \widehat{\xi}_{n,n}^i)$  taking some value  $(x_0, \dots, x_n)$  is given by

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N 1_{(x_0, \dots, x_n)}(\widehat{\xi}_{0,n}^i, \widehat{\xi}_{1,n}^i, \dots, \widehat{\xi}_{n,n}^i) = \frac{1}{\mathcal{Z}_{n+1}} \exp \left( - \sum_{0 \leq p \leq n} \beta_p V(x_p) \right) \mathbb{P}_n(x_0, \dots, x_n)$$

with the probability of a free evolution path involving only mutation transitions

$$\mathbb{P}_n(x_0, \dots, x_n) = \eta_0(x_0) M_1(x_0, x_1) \dots M_n(x_{n-1}, x_n)$$

and  $\mathcal{Z}_{n+1}$  is a normalizing constant.

Suppose that every free evolution path has the same chance to be sampled, in the sense that

$$\mathbb{P}_n(x_0, \dots, x_n) = \mathbb{P}_n(y_0, \dots, y_n)$$

for any paths admissible pair of paths  $(x_0, \dots, x_n)$  and  $(y_0, \dots, y_n)$ . This condition is satisfied if  $\eta_0$  is the uniform counting measure on  $E$  and the mutation transitions  $M_n(x, y)$  correspond to local random choices of the same number of neighbors, starting from any state  $x$ . In this case, for any admissible path  $(x_0, \dots, x_n)$  we have that

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N 1_{(x_0, \dots, x_n)}(\widehat{\xi}_{0,n}^i, \widehat{\xi}_{1,n}^i, \dots, \widehat{\xi}_{n,n}^i) = \frac{1}{\mathcal{Z}'_n} \exp \left( - \sum_{0 \leq p \leq n} \beta_p V(x_p) \right) \quad (1.16)$$

for some normalizing constant  $\mathcal{Z}'_n$ . When the inverse temperature parameter  $\beta_p$  increases the r.h.s. probability mass quantity only charges admissible paths  $(x_0, \dots, x_n)$  that minimize the path potential function

$$\mathcal{V}_n(x_0, \dots, x_n) = \inf_{(y_0, \dots, y_n)} \sum_{0 \leq p \leq n} V(y_p)$$

In other words at low temperature, the ancestral lines of the simple genetic model described above converge to the uniform measure on all the paths  $(x_0, \dots, x_n)$  of length  $n$  that minimize the energy function  $\mathcal{V}_n$ . For time homogenous mutation transitions associated with stochastic matrices  $M_n(x, y) = M(x, y)$  satisfying the following condition for some integer  $m \geq 1$  and any pair  $(x, y) \in E^2$

$$M(x, x) > 0 \quad \text{and} \quad M^m(x, y) \geq \epsilon M^m(x, z)$$

we also have the convergence result

$$\lim_{n \rightarrow \infty} \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N 1_{V^*}(\widehat{\xi}_n^i) = 1$$

as soon as  $\beta_n = C \log(n+1)$  for some constant  $C$  that depends on  $m$  and on the oscillations of the function  $V$ . This convergence result is also true for  $\beta_n = C(n+1)^\alpha$ , with any  $\alpha \in ]0, 1[$ , as soon as the above condition is met for  $m = 1$ . Further details on these concentration properties can be found in [14]. Related convergence results for fixed population sizes can be found in [4]. To give a flavor of these results, let us suppose that the mutation transitions  $M_n(x, y)$  also depend on the inverse temperature parameter and

$$M_n(x, y) \rightarrow_{n \rightarrow \infty} 1_x(y) \quad \text{as} \quad \beta_n \uparrow \infty$$

Intuitively speaking, the genetic mutations become rare transitions at low temperature. In this situation, we can prove that there exists a “critical population size”  $N^*$  that depends on the energy function as well as on the free evolution model such that

$$\forall N \geq N^* \quad \lim_{n \rightarrow \infty} \text{Proba} \left( \forall 1 \leq i \leq N \quad \widehat{\xi}_n^i \in V^* \right) = 1$$

The particle algorithm described above is more general than the interacting simulated annealing models discussed in section 1.4. If we replace the temperature parameters  $\beta_n$  by temperature increments  $(\beta_n - \beta_{n-1})$ , and if we choose the MCMC type mutation transitions  $M_{\beta_n}$  discussed in section 1.4, then this model coincides with the one presented in 1.4. In more general situations, the limiting Feynman-Kac type measures (1.16) differ from the Boltzmann-Gibbs measures (1.12) and (1.14) presented in section 1.4 and section 1.5.

### 1.7. Directed polymers and self avoiding walks

In biology and chemistry, flexible polymers describe the kinetic structure of macromolecules in a given chemical solvent. The polymer chain at time  $n$  is regarded as a sequence of random variables

$$X_n = (X'_0, \dots, X'_n) \in E_n = \underbrace{E \times \dots \times E}_{(n+1) \text{ times}}$$

taking values in some finite or countable state space  $E$ . Of course, in concrete situations the state space is not a finite set but some Euclidian state space  $E = \mathbb{R}^d$ , for some  $d \geq 1$ . This countable state space condition is only made to avoid unnecessary technicalities in the presentation of these stochastic models in path spaces. The following discussion is mainly taken from section 12.5 in the book [9].

The elementary states  $X'_p$  represent the different directed monomers of the macromolecules  $X_n$ . The length parameter  $n$  represents the degree of polymerization. The monomers are connected by chemical bonds and interact with one another as well as with the chemicals in the solvent. The energy of a polymerization sequence

$$X_0 = X'_0 \longrightarrow X_1 = (X'_0, X'_1) \longrightarrow X_2 = (X'_0, X'_1, X'_2) \longrightarrow \dots \longrightarrow X_n = (X'_0, \dots, X'_n)$$

is defined in terms of a Boltzmann potential

$$\exp\left(-\beta \sum_{p=0}^n V_p(X'_0, \dots, X'_p)\right) \quad (1.17)$$

The parameter  $\beta \in \mathbb{R}_+$  represents the inverse temperature of the solvent, and each potential function

$$V_n : (x_0, \dots, x_n) \in E_n \rightarrow V_n(x_0, \dots, x_n) \in \mathbb{R}_+$$

reflects the local intermolecular energy between the monomer  $X'_n = x_n$  in the polymer chain  $X_{n-1} = (x_0, \dots, x_{n-1})$  during the  $n$ th polymerization

$$X_{n-1} = (x_0, \dots, x_{n-1}) \longrightarrow X_n = ((x_0, \dots, x_{n-1}), x_n)$$

The potential functions  $V_n$  depend on the nature of the solvent and the physico-chemical structure of the polymer. At low temperature,  $\beta \uparrow \infty$ , the interaction between monomers may be strongly repulsive at short distances and attractive or repulsive at larger ones. For instance, the monomers may tend to avoid being closed on each other. These excluded volume effects and repulsive interactions can be modeled by choosing a potential function satisfying the following condition:

$$V_n(x_0, \dots, x_n) = 0 \iff x_n \notin \{x_0, \dots, x_{n-1}\} \quad (1.18)$$

In this situation, every self-interaction is penalized by a factor  $e^{-\beta V_n}$  so that the energy of an elementary polymerization is minimal if and only if the new monomer differs from the previous ones. In this context, the inverse temperature parameter  $\beta$  is sometimes called the strength of repulsion.

In the opposite case, at high temperature,  $\beta \rightarrow 0$ , the interaction forces disappear. In this situation, it is commonly assumed that  $X'_n$  is an  $E$ -valued Markov chain with elementary transitions  $M'_n$  and initial

distribution  $\eta'_0$ . By the definition of the chain  $X_n = (X'_0, \dots, X'_n)$ , this Markovian hypothesis implies that the Markov transitions  $M_n$  have the form

$$M_n((x_0, \dots, x_{n-1}), (y_0, \dots, y_{n-1}, y_n)) = 1_{(x_0, \dots, x_{n-1})}(y_0, \dots, y_{n-1}) M'_n(y_{n-1}, dy_n)$$

for any paths  $(x_0, \dots, x_{n-1}) \in E_{n-1}$  and  $(y_0, \dots, y_n) \in E_n$  and for some Markov transition  $M'_n$  from  $E$  into  $E$ .

The Boltzmann-Gibbs measures associated with these directed polymer models are the measures in the product spaces  $E_n$  defined below

$$\mathbb{Q}_n(x_0, \dots, x_n) := \frac{1}{\mathcal{Z}_n} \exp\left(-\beta \sum_{p=0}^{n-1} V_p(x_0, \dots, x_p)\right) \mathbb{P}_n(x_0, \dots, x_n)$$

with the distribution of the free paths of the monomer chain

$$\mathbb{P}_n(x_0, \dots, x_n) = \eta'_0(x_0) M'_1(x_0, x_1) \dots M'_n(x_{n-1}, x_n)$$

These measures can alternatively be rewritten in the following form

$$\mathbb{Q}_n(x_0, \dots, x_n) := \frac{1}{\mathcal{Z}_n} \left\{ \prod_{p=0}^{n-1} G_p(x_0, \dots, x_p) \right\} \mathbb{P}_n(x_0, \dots, x_n)$$

with the exponential weight functions

$$\forall n \geq 0 \quad G_n(x_0, \dots, x_n) := \exp(-\beta V_n(x_0, \dots, x_n))$$

To illustrate these models, let us suppose that  $X'_n$  is a simple random walk on the  $d$ -dimensional lattice  $\mathbb{Z}^d$  starting at the origin and set

$$G_n(x_0, \dots, x_n) := 1_{\mathbb{Z}^d - \{x_0, \dots, x_{n-1}\}}(x_n)$$

In this situation, we have that

$$\mathbb{Q}_n = \text{Law}(X'_0, \dots, X'_n \mid \forall 0 \leq p < q < n \ X'_p \neq X'_q)$$

and

$$\mathcal{Z}_n = \text{Proba}(\forall 0 \leq p < q < n \ X'_p \neq X'_q) = \text{Card}(\mathbf{S}_{n-1}) / (2d)^{n-1}$$

where  $\mathbf{S}_{n-1}$  is the set of self-avoiding walks of length  $n-1$  starting at the origin.

The  $N$  particle approximation of these quantities is nothing but a simple genetic type evolution model with  $N$  path-valued particles

$$\begin{aligned} \xi_n^i &:= (\xi_{0,n}^i, \xi_{1,n}^i, \dots, \xi_{n,n}^i) \\ \widehat{\xi}_n^i &:= (\widehat{\xi}_{0,n}^i, \widehat{\xi}_{1,n}^i, \dots, \widehat{\xi}_{n,n}^i) \in E_n \end{aligned}$$

During the selection stage, with a probability  $G_n(\xi_n^i)$  every path-valued individual stays in the same place  $\widehat{\xi}_n^i = \xi_n^i$ ; otherwise, we replace  $\xi_n^i$  by a new individual  $\widehat{\xi}_n^i = \xi_n^j$  randomly chosen among the individuals  $\xi_0^j$  with a probability proportional to its weight  $G_n(\xi_n^i)$ . This mechanism is intended to favor minimal energy polymerizations. For instance, in the case of repulsive interaction (1.18), a given polymer with degree  $n$ , say  $(\xi_{0,n}^i, \dots, \xi_{n,n}^i)$ , has more chance of being selected if the last monomer  $\xi_{n,n}^i$  added during the  $n$ th sampled polymerization differs from the previous ones; that is, if  $\xi_{n,n}^i \notin \{\xi_{0,n}^i, \dots, \xi_{n-1,n}^i\}$ . Each selected polymer is

a path-valued individual  $\widehat{\xi}_n^i$ . During the mutation transition,  $\widehat{\xi}_n^i$  evolves randomly according to the transition  $M_{n+1}$  of the path-valued Markov chain  $X_n$ .

$$\begin{aligned}\xi_{n+1}^i &= ((\xi_{0,n+1}^i, \dots, \xi_{n,n+1}^i), \xi_{n+1,n+1}^i) \\ &= ((\widehat{\xi}_{0,n}^i, \dots, \widehat{\xi}_{n,n}^i), \xi_{n+1,n+1}^i) \in E_{n+1} = (E_n \times E)\end{aligned}\quad (1.19)$$

where  $\xi_{n+1,n+1}^i$  is a random variable with distribution  $M'_n(\widehat{\xi}_n^i, \cdot)$ . Various asymptotic estimates can be derived. For instance, for any function  $f$  on  $E_n = E^{(n+1)}$  and any time horizon  $n$ , we have

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N f(\xi_{0,n}^i, \xi_{1,n}^i, \dots, \xi_{n,n}^i) = \sum_{x_0, \dots, x_n} \mathbb{Q}_n(x_0, \dots, x_n) f(x_0, \dots, x_n)$$

and unbiased estimates of the normalizing constants

$$\mathcal{Z}_n^N := \prod_{0 \leq p < n} \eta_p^N(G_p) \xrightarrow{N \rightarrow \infty} \mathcal{Z}_n$$

where  $\eta_p^N(G_p)$  is the empirical mean value of the potential functions after the mutation stage

$$\forall n \geq 0 \quad \eta_n^N(G_n) := \frac{1}{N} \sum_{1 \leq i \leq N} G_n(\xi_{0,n}^i, \xi_{1,n}^i, \dots, \xi_{n,n}^i)$$

## 2. A BRIEF TREATISE ON FEYNMAN-KAC MODELING

### 2.1. Stochastic matrices and Markov semigroups

We let  $E$  be a finite set. A Markov chain  $X = (X_n)_{n \geq 0}$  is defined by a matrix  $M$  with positive entries  $M(x_1, x_2) \geq 0$  s.t.  $\sum_{x_2} M(x_1, x_2) = 1$  for each state  $x_1$ . Such matrices are called transition matrices or stochastic matrices in the literature on Markov chains. For instance, for finite state spaces  $E = \{1, \dots, d\}$ , with cardinality  $d \geq 1$ , we can identify a Markov transition  $M(x, y)$  from  $E$  into itself with a  $d \times d$  matrix

$$\begin{pmatrix} M(1,1) & \cdots & M(1,d) \\ \vdots & \vdots & \vdots \\ M(d,1) & \cdots & M(d,d) \end{pmatrix}.$$

Starting from  $x_1$  we sample the chain transition by choosing randomly a state  $x_2$  with distribution  $M(x_1, x_2)$ . The outcomes  $X_0 = x_0, X_1 = x_1, X_2 = x_2, \dots$  are referred as a realization of the chain  $X$  starting from  $X_0 = x_0$ . By construction, we have

$$\mathbb{P}(X_n = x_n \mid X_{n-1} = x_{n-1}) = M(x_{n-1}, x_n)$$

and more generally

$$\mathbb{P}(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n \mid X_0 = x_0) = M(x_0, x_1)M(x_1, x_2) \dots M(x_{n-1}, x_n)$$

These simple observations already show that the law of the random state  $X_n$  at time  $n$  starting from  $X_0 = x_0$  is given by the  $n$ -th power of the matrix  $M$ ; that is we have

$$\mathbb{P}(X_n = x_n \mid X_0 = x_0) = \sum_{x_1, \dots, x_{n-1}} M(x_0, x_1)M(x_1, x_2) \dots M(x_{n-1}, x_n) = M^n(x_0, x_n)$$



Identifying functions  $f : x \in E \mapsto f(x) \in \mathbb{R}$  with column type vector  $f = (f(x))_{x \in E}$ , the conditional expectations of the random variable  $f(X_n)$  given  $X_0 = x_0$  are also given by a matrix operation

$$\mathbb{E}(f(X_n) | X_0 = x_0) = \sum_{x_n} M^n(x_0, x_n) f(x_n) = M^n(f)(x_0)$$

where  $M^n(f)$  stands for the function given by the column vector  $M^n(f) = (M^n(f)(x_0))_{x_0 \in E}$ . For instance, for finite state spaces  $E = \{1, \dots, d\}$ , functions

$$f(x) = \sum_{i=1}^d f(i) 1_i(x)$$

are identified with the column vector

$$f = (f(1), \dots, f(d))^* := \begin{pmatrix} f(1) \\ \vdots \\ f(d) \end{pmatrix}. \quad (2.1)$$

In this situation, the function  $M(f)$  defined by

$$M(f)(x) := \mathbb{E}(f(X_n) | X_{n-1} = x)$$

is given by the finite sum

$$\forall 1 \leq i \leq d \quad M(f)(i) = \sum_{j=1}^d M(i, j) f(j).$$

Using the vector notation we find that the function  $M(f)$  can alternatively be defined by the following matrix operation:

$$M(f) := \begin{pmatrix} M(f)(1) \\ \vdots \\ M(f)(d) \end{pmatrix} = \begin{pmatrix} M(1, 1) & \cdots & M(1, d) \\ \vdots & \vdots & \vdots \\ M(d, 1) & \cdots & M(d, d) \end{pmatrix} \begin{pmatrix} f(1) \\ \vdots \\ f(d) \end{pmatrix} = Mf$$

In the same way, identifying probability measures  $\eta_0$  on  $E$  with row type vectors  $\eta_0 = (\eta_0(x_0))_{x_0 \in E}$  s.t.  $\eta_0(x_0) \geq 0$  and  $\sum_{x_0} \eta_0(x_0) = 1$ , the distribution  $\eta_n$  of the state  $X_n$  of the chain starting randomly from  $X_0$  with distribution  $\eta_0$  is given by

$$\eta_n(x_n) = \mathbb{P}(X_n = x_n) = \sum_{x_0} \eta_0(x_0) M^n(x_0, x_n) = \eta_0 M^n(x_n)$$

where  $\eta_0 M^n$  stands for the probability on  $E$  given by the row vector  $\eta_0 M^n = (\eta_0 M^n(x_n))_{x_n \in E}$ . Furthermore, we have

$$\eta_n(f) = \sum_{x_n} \eta_n(x_n) f(x_n) = \mathbb{E}(f(X_n))$$

In matrix notation, we simply have that

$$\eta_n = \eta_0 M^n \quad \text{and} \quad \eta_n(f) = (\eta_0 M^n)(f) = \eta_0([M^n(f)])$$

For instance, for instance, for finite state spaces  $E = \{1, \dots, d\}$ , the sum  $\eta_n(f)$  can be rewritten as

$$\eta_n(f) = \eta_n f = (\eta_n(1), \dots, \eta_n(d)) \begin{pmatrix} f(1) \\ \vdots \\ f(d) \end{pmatrix} = \sum_{i=1}^d f(i) \eta_n(i).$$

and we have

$$\begin{aligned} (\eta_0 M) &:= [(\eta_0 M)(1), \dots, (\eta_0 M)(d)] \\ &= [\eta_0(1), \dots, \eta_0(d)] \begin{pmatrix} M(1,1) & \cdots & M(1,d) \\ \vdots & \ddots & \vdots \\ M(d,1) & \cdots & M(d,d) \end{pmatrix} = \eta_0 M. \end{aligned}$$

In much the same way, we find that

$$\eta_n = \eta_{n-1} M = (\eta_{n-2} M) M = \eta_{n-2} M^2 = \dots = \eta_0 M^n$$

By construction, the law of the random states  $X_p$  and  $X_n$ , with  $0 \leq p \leq n$  are given by the following matrix operation

$$\eta_n = \eta_p M^{(n-p)}$$

The collection of matrices  $M^n$ , with  $n \geq 0$ , is called the Markov semigroup associated with the transition probabilities  $M$ . For  $n = 0$ , we use the convention  $M^0 = Id$  the identity matrix with zero off-diagonal elements and unit diagonal elements.

For instance, if  $E = \{1, 2\}$  then the Markov transition  $M$  is given by the  $2 \times 2$ -matrix

$$M = \begin{pmatrix} M(1,1) & M(1,2) \\ M(2,1) & M(2,2) \end{pmatrix}$$

We further assume that the entries  $M(i, j) \in [0, 1]$  are chosen s.t.  $c := M(1, 2) + M(2, 1) > 0$ . Using an elementary induction on the time parameter  $n$ , we prove

$$M^n = \frac{1}{c} \begin{pmatrix} M(2,1) & M(1,2) \\ M(2,1) & M(1,2) \end{pmatrix} + \frac{(1-c)^n}{c} \begin{pmatrix} M(1,2) & -M(1,2) \\ -M(2,1) & M(2,1) \end{pmatrix}.$$

Except in some very particular cases, the matrix semigroup  $M^n$  has no explicit expression.

Time nonhomogenous models can be studied in the same line of arguments. For non homogeneous Markov transitions  $M_n$ , we have

$$\mathbb{P}(X_n = x_n \mid X_{n-1} = x_{n-1}) = M_n(x_{n-1}, x_n)$$

The law of the chain  $X_n$  is now given by the following matrix composition

$$\eta_n = \eta_0 M_1 M_2 \dots M_n$$

The semigroup  $M_{p,n}$  associated with this sequence of measures is now given by the following formulae

$$\eta_n = \eta_p M_{p,n} \quad \text{with the matrix} \quad M_{p,n} := M_{p+1} M_{p+2} \dots M_n$$

## 2.2. The Bayes-Boltzmann-Gibbs transformation

We consider a non negative potential function  $G$  and a probability measure  $\eta$  on a finite or countable state space  $E$ . We recall that  $\eta(G)$  is the quantity defined by the sum

$$\eta(G) := \sum_x \eta(x) G(x)$$

For indicator potential functions  $G = I_A$ , sometimes we slightly abuse notation and we set

$$\eta(1_A) = \sum_x \eta(x) 1_A(x) = \sum_{x \in A} \eta(x) := \eta(A)$$

**Definition 2.1.** We let  $\Psi_G(\eta)$  be the probability measure on  $E$  defined by

$$\Psi_G(\eta)(x) := \frac{1}{\eta(G)} G(x) \eta(x)$$

Note that this probability measure  $\Psi_G(\eta)$  is well-defined if and only if  $\eta(G) > 0$ . The transformation  $\Psi_G$  is called the *Bayes-Boltzmann-Gibbs transformation* (abbreviate BBG-transformation) associated to the potential function  $G$ .

When  $G = 1_A$ , we notice that  $\Psi_G(\eta)$  is the restriction of the measure  $\eta$  to the set  $A$ ;

$$\Psi_{1_A}(\eta)(x) := \eta_A(x) := \frac{1}{\eta(A)} 1_A(x) \eta(x).$$

In this case, we have

$$A \subset B \implies \Psi_{1_A}(\Psi_{1_B}(\eta)) = \Psi_{1_B}(\Psi_{1_A}(\eta)) = \eta_{A \cap B} = \Psi_{1_{A \cap B}}(\eta) = \Psi_{1_A}(\eta)$$

In nonlinear filtering problems discussed in section 1.3, the measure  $\eta$  represents the conditional law of a Markov chain  $X_n$  given a series of observations  $(Y_0, \dots, Y_{n-1})$  and  $G$  is the likelihood function  $G(x_n) = g_n(y_n, x_n)$  associated with an observation  $Y_n = y_n$  delivered by some sensor. In this context the BBG transformation describes the way a new observation is incorporated into the filtering distribution

$$\eta = \text{Law}(X_n \mid (Y_0, \dots, Y_{n-1})) \rightsquigarrow \Psi_{g_n(y_n, \cdot)}(\eta) = \text{Law}(X_n \mid (Y_0, \dots, Y_{n-1}, \mathbf{Y}_n))$$

We shall return more precisely to this BBG transformation in the further developments of the article. In matrix notation, if the state space is given by  $E = \{1, \dots, d\}$  we can also identify the multiplicative potential function  $G$  with the following diagonal matrix

$$\mathcal{G} = \begin{pmatrix} G(1) & 0 & \cdots & 0 & 0 \\ 0 & G(2) & 0 \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & G(d-1) & 0 \\ 0 & 0 & \cdots & 0 & G(d) \end{pmatrix}.$$

Using the row and column vector identifications, we find that

$$\eta \mathcal{G} = (\eta(1), \dots, \eta(d)) \mathcal{G} = (\eta(1)G(1), \dots, \eta(d)G(d))$$

with the normalizing constant given by the  $\mathbb{L}_1$ -norm of the vector  $\eta \mathcal{G}$

$$\eta(G) = \|\eta \mathcal{G}\|_1 = \sum_{i=1}^d (\eta \mathcal{G})(i) = \eta \mathcal{G}(1) = \eta \mathcal{G} \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} = \sum_{i=1}^d \mu(x_i) G(x_i).$$

In this matrix notation, the BBG transformation takes the following form

$$\Psi_G(\eta) = \frac{1}{\|\eta \mathcal{G}\|_1} \eta \mathcal{G}.$$

We end this section with a nonlinear Markov interpretation of the BBG-transformation.

**Proposition 2.2.** *For any collection of non negative parameters  $\epsilon(\eta, G)$  that may depend on  $G$  and on  $\eta$ , and such that*

$$\epsilon(\eta, G) G(x) \leq 1 \quad \text{for every } x \in E \quad \text{s.t. } \eta(x) > 0$$

*we have the following decomposition*

$$\forall y \in E \quad \Psi_G(\eta)(y) = (\eta S_\eta)(y) := \sum_x \eta(x) S_\eta(x, y) \quad (2.2)$$

*with the collection of selection type Markov transitions on  $E$  defined by*

$$S_\eta(x, y) = \epsilon(\eta, G) G(x) 1_x(y) + (1 - \epsilon(\eta, G) G(x)) \Psi_G(\eta)(y).$$

Before getting into the details of the proof of this proposition, for finite state spaces  $E = \{1, \dots, d\}$ , the BBG-transformation takes the following matrix form

$$\begin{aligned} (\eta S_\eta) &:= [(\eta S_\eta)(1), \dots, (\eta S_\eta)(d)] \\ &= [\eta(1), \dots, \eta(d)] \begin{pmatrix} S_\eta(1, 1) & \cdots & S_\eta(1, d) \\ \vdots & \vdots & \vdots \\ S_\eta(d, 1) & \cdots & S_\eta(d, d) \end{pmatrix} = \eta S_\eta = \Psi_G(\eta). \end{aligned}$$

**Proof:**

To check this assertion, we choose a function  $f$  on  $E$  and observe that

$$\begin{aligned} S_\eta(f)(x) &:= \sum_y S_\eta(x, y) f(y) \\ &= \epsilon(\eta, G) G(x) f(x) + (1 - \epsilon(\eta, G) G(x)) \Psi_G(\eta)(f) \end{aligned}$$

from which we find that

$$\begin{aligned} \eta(S_\eta(f)) &:= \sum_x \eta(x) S_\eta(f)(x) \\ &= \epsilon(\eta, G) \eta(Gf) + (1 - \epsilon(\eta, G) \eta(G)) \Psi_G(\eta)(f) \\ &= \epsilon(\eta, G) \eta(Gf) + \Psi_G(\eta)(f) - \epsilon(\eta, G) \eta(G) \frac{\eta(Gf)}{\eta(G)} = \Psi_G(\eta)(f). \end{aligned}$$

This ends the proof of the proposition. ■

### 2.3. Positive matrices and Feynman-Kac measures

Let  $E$  be a finite set. We consider a collection of matrices  $Q_n := (Q_n(x, y))_{x, y \in E}$  with non negative entries  $Q_n(x, y) \geq 0$ . To simplify the presentation, we further assume that the entries are strictly positive. Given a probability measure  $\eta_0$  on  $E$ , we denote by  $\Gamma_n$  the measure on  $E^{(n+1)}$  defined for any path  $(x_p)_{0 \leq p \leq n} \in E^{(n+1)}$  of length  $n$  by the following formula:

$$\Gamma_n(x_0, \dots, x_n) := \eta_0(x_0) Q_1(x_0, x_1) Q_2(x_1, x_2) \cdots Q_n(x_{n-1}, x_n)$$

Notice that  $\Gamma_n$  can alternatively be defined by the following recursive formula

$$\forall n \geq 1 \quad \Gamma_n(x_0, \dots, x_n) = \Gamma_{n-1}(x_0, \dots, x_{n-1}) \times Q_n(x_{n-1}, x_n)$$

with  $\Gamma_0(x_0) = \eta_0(x_0)$ . We choose a function  $f_n$  on path space  $E^{(n+1)}$  and we set

$$\Gamma_n(f_n) = \sum_{x_0, \dots, x_n} \Gamma_n(x_0, \dots, x_n) f_n(x_0, \dots, x_n) \quad (2.3)$$

Feynman-Kac models provide a precise meaning of these sums. To be more precise, we set

$$M_n(x, y) := \frac{Q_n(x, y)}{\sum_z Q_n(x, z)} \quad \text{and} \quad G_n(x) := \sum_z Q_{n+1}(x, z) = Q_n(1)(x)$$

The r.h.s. term in the above display expresses the sum in terms of a matrix type operation with the unit function 1, defined by  $1(x) = 1$  for any  $x \in E$ . By construction, it is readily checked that

$$\begin{aligned} \Gamma_n(x_0, \dots, x_n) &:= \eta_0(x_0) \left\{ \prod_{1 \leq p \leq n} Q_p(1)(x_{p-1}) \right\} \times \left\{ \prod_{1 \leq p \leq n} \frac{Q_p(x_{p-1}, x_p)}{Q_p(1)(x_{p-1})} \right\} \\ &= \left\{ \prod_{0 \leq p < n} G_p(x_p) \right\} \times \left\{ \eta_0(x_0) \prod_{1 \leq p \leq n} M_p(x_{p-1}, x_p) \right\} \end{aligned}$$

We conclude that

$$\Gamma_n(x_0, \dots, x_n) = \left\{ \prod_{0 \leq p \leq n} G_p(x_p) \right\} \times \mathbb{P}_n(x_0, \dots, x_n)$$

with the law of the path  $(x_p)_{0 \leq p \leq n}$  of a non homogeneous Markov chain  $X_n$  with initial distribution  $\eta_0 = \text{Law}(X_0)$  and local Markov transitions  $M_n$

$$\mathbb{P}_n(x_0, \dots, x_n) := \eta_0(x_0) \prod_{1 \leq p \leq n} M_p(x_{p-1}, x_p)$$

In this interpretation, the quantities (2.3) can alternatively be rewritten in terms of a Feynman-Kac formula

$$\Gamma_n(f_n) = \mathbb{E} \left( f_n(X_0, \dots, X_n) \prod_{0 \leq p < n} G_p(X_p) \right)$$

It is also important to observe that

$$\Gamma_n(1) = \sum_{x_0, \dots, x_n} \Gamma_n(x_0, \dots, x_n) = \mathbb{E} \left( \prod_{0 \leq p < n} G_p(X_p) \right)$$

The above constructions clearly works for any decomposition of  $Q_n$  in the following form

$$Q_n(x, y) = G_{n-1}(x) M_n(x, y)$$

for some Markov transition  $M_n$  and some potential function  $G_{n-1}$ . Note that in this case we have

$$Q_n(1)(x) = \sum_y Q_n(x, y) = G_{n-1}(x) \sum_y M_n(x, y) = G_{n-1}(x)$$

**Definition 2.3.** We let  $X_n$  be a Markov chain with Markov transitions  $M_n$  on some finite or countable state space  $E$ , and we consider a sequence of  $[0, 1]$ -valued potential functions  $G_n$  on the set  $E$ . The Feynman-Kac path measure associated to the pairs  $(M_n, G_n)$  is the measure  $\mathbb{Q}_n$  defined for any  $f_n$  on  $E^{(n+1)}$  by the following formulae

$$\mathbb{Q}_n(f_n) := \frac{\Gamma_n(f_n)}{\Gamma_n(1)} \quad \text{with} \quad \Gamma_n(f_n) = \mathbb{E} \left( f_n(X_0, \dots, X_n) \prod_{0 \leq p < n} G_p(X_p) \right) \quad (2.4)$$

Notice that  $\mathbb{Q}_n$  can alternatively be defined by the following formulation:

$$\mathbb{Q}_n(x_0, \dots, x_n) = \frac{1}{\mathcal{Z}_n} \left\{ \prod_{0 \leq p < n} G_p(x_p) \right\} \mathbb{P}_n(x_0, \dots, x_n)$$

with the normalizing constant  $\mathcal{Z}_n := \Gamma_n(1)$ , and the distribution of the free path evolution of the reference Markov chain  $X_n$  given by:

$$\mathbb{P}_n(x_0, \dots, x_n) = \eta_0(x_0) \prod_{1 \leq p \leq n} M_p(x_{p-1}, x_p)$$

We also consider the measure  $\widehat{\mathbb{Q}}_n$  on  $E^{(n+1)}$  defined as  $\mathbb{Q}_n$  by taking a product of potential functions from the origin  $p = 0$ , up to the final time horizon  $p = n$

$$\widehat{\mathbb{Q}}_n(x_0, \dots, x_n) = \frac{1}{\widehat{\mathcal{Z}}_n} \left\{ \prod_{0 \leq p \leq n} G_p(x_p) \right\} \mathbb{P}_n(x_0, \dots, x_n)$$

with the normalizing constant  $\widehat{\mathcal{Z}}_n = \mathcal{Z}_{n+1}$ .

Another important quantities are the  $n$ -th marginals  $\eta_n$  and  $\widehat{\eta}_n$  of the measure  $\mathbb{Q}_n$  and  $\widehat{\mathbb{Q}}_n$  defined by

$$\eta_n(x_n) = \sum_{x_0, \dots, x_{n-1}} \mathbb{Q}_n(x_0, \dots, x_{n-1}, x_n) \quad \text{and} \quad \widehat{\eta}_n(x_n) = \sum_{x_0, \dots, x_{n-1}} \widehat{\mathbb{Q}}_n(x_0, \dots, x_{n-1}, x_n) \quad (2.5)$$

It is important to observe that this pair of measures can be alternatively defined for any function  $f$  on  $E$  by the following formulae

$$\eta_n(f) := \gamma_n(f)/\gamma_n(1) \quad \text{with} \quad \gamma_n(f) = \mathbb{E} \left( f(X_n) \prod_{0 \leq p < n} G_p(X_p) \right) \quad (2.6)$$

and

$$\widehat{\eta}_n(f) = \widehat{\gamma}_n(f)/\widehat{\gamma}_n(1) \quad \text{with} \quad \widehat{\gamma}_n(f) = \mathbb{E} \left( f(X_n) \prod_{0 \leq p \leq n} G_p(X_p) \right) \quad (2.7)$$

For  $n = 0$ , we use the convention  $\prod_{\emptyset} = 1$ , so that  $\eta_0 = \text{Law}(X_0)$ . The above assertions can be proved directly from the definitions of the measures  $\mathbb{Q}_n$  and  $\widehat{\mathbb{Q}}_n$ . In reference to filtering, the measure  $\eta_n$  is often called the prediction measure at time  $n$ , while  $\widehat{\eta}_n$  is referred as the updated measure at time  $n$ .

We end this section with the following elementary observation connecting the pair of measures  $\mathbb{Q}_n$  and  $\widehat{\mathbb{Q}}_n$ .

$$\begin{aligned} \left\{ \prod_{0 \leq p \leq n} G_p(x_p) \right\} \mathbb{P}_n(x_0, \dots, x_n) &= \eta_0(x_0) G_0(x_0) \prod_{1 \leq p \leq n} M_p(x_{p-1}, x_p) G_p(x_p) \\ &= \eta_0(G_0) \left\{ \prod_{0 \leq p < n} \widehat{G}_p(x_p) \right\} \widehat{\mathbb{P}}_n(x_0, \dots, x_n) \end{aligned}$$

with the probability measure  $\widehat{\mathbb{P}}_n$  defined below

$$\widehat{\mathbb{P}}_n(x_0, \dots, x_n) := \widehat{\eta}_0(x_0) \prod_{1 \leq p \leq n} \widehat{M}_p(x_{p-1}, x_p)$$

and the pair  $(\widehat{G}_p, \widehat{M}_p)$  of potential and transitions given by

$$\widehat{M}_p(x_{p-1}, x_p) = \frac{M_p(x_{p-1}, x_p) G_p(x_p)}{M_p(G_p)(x_{p-1})} \quad \text{and} \quad \widehat{G}_p(x_p) = M_{p+1}(G_{p+1})(x_p)$$

These formulae imply that

$$\widehat{\mathbb{Q}}_n(x_0, \dots, x_n) = \frac{1}{\widehat{\mathcal{Z}}'_n} \left\{ \prod_{0 \leq p < n} \widehat{G}_p(x_p) \right\} \widehat{\mathbb{P}}_n(x_0, \dots, x_n)$$

with some normalizing constant  $\widehat{\mathcal{Z}}'_n$ . In other words  $\widehat{\mathbb{Q}}_n$  coincides with the measure defined as  $\mathbb{Q}_n$  by replacing the pair  $(G_n, M_n)$  by the pair of updated type quantities  $(\widehat{G}_n, \widehat{M}_n)$ . From the pure mathematical point of view, this rather elementary observation shows that the two models  $\mathbb{Q}_n$  and  $\widehat{\mathbb{Q}}_n$  are equivalent.

#### 2.4. Examples as Feynman-Kac models

We illustrate the rather abstract models presented in section 2.3 in the examples discussed in section 1.

##### (1) Random walks confined to a set

We let  $M_n(x, y)$  be the transitions of a symmetric random walk  $X_n$  on the integers  $\mathbb{Z}$  starting at the origin  $X_0 = 0$ . We also fix a interval  $A = \{-a + 1, -a + 2, \dots, a - 1\}$ , with  $a > 0$ , and we set  $G_n := 1_A$ . In this situation, we have

$$\mathbb{Q}_n = \text{Law}((X_0, \dots, X_n) \mid \forall 0 \leq p < n, X_p \in A)$$

and

$$\widehat{\mathbb{Q}}_n = \text{Law}((X_0, \dots, X_n) \mid \forall 0 \leq p \leq n, X_p \in A)$$

##### (2) Particle absorption models

For any  $[0, 1]$ -valued potential functions  $G_n$  and any Markov transitions  $M_n$ , the measures  $\mathbb{Q}_n$  have the following interpretations

$$\mathbb{Q}_n = \text{Law}((X_0^c, \dots, X_n^c) \mid T \geq n)$$

and

$$\widehat{\mathbb{Q}}_n = \text{Law}((X_0^c, \dots, X_n^c) \mid T > n)$$

where  $X_n^c$  is the particle absorption model defined in (1.8), and  $T$  the random absorption time of the particle with  $M_n$ -free evolution and absorption rates  $G_n$ .

(3) **Nonlinear filtering problems**

A given Markov chain  $X_n$  with transitions  $M_n$  is partially observed at every time step  $n$  by some noisy sensor that delivers a random observation  $Y_n = y_n$  with distribution

$$\text{Proba}(Y_n = y_n \mid X_n = x_n) = G(x_n, y_n)$$

For every fixed sequence of observations  $Y_n = y_n$ , we set  $G_n(x) = G(x, y_n)$ . In this situation, the measures  $\mathbb{Q}_n$  and their normalizing constants  $\mathcal{Z}_n$  have the following interpretations

$$\mathbb{Q}_n = \text{Law}((X_0, \dots, X_n) \mid \forall 0 \leq p < n, Y_p = y_p)$$

and

$$\mathcal{Z}_n = \mathbb{P}((Y_0, \dots, Y_{n-1}) = (y_0, \dots, y_{n-1}))$$

In the same way, we have

$$\widehat{\mathbb{Q}}_n = \text{Law}((X_0, \dots, X_n) \mid \forall 0 \leq p \leq n, Y_p = y_p)$$

(4) **Stochastic optimization algorithms**

We consider a collection of MCMC transitions  $M_\beta(x, y)$ , indexed by some inverse temperature parameter  $\beta \geq 0$ , on some finite state space  $E$  equipped with the counting measure  $\lambda(x) := 1/\text{Card}(E)$ . We assume that  $M_\beta(x, y)$  is a Markov transition with an invariant Boltzmann-Gibbs distribution  $\mu_\beta$  associated with a non negative cost function  $V$  on  $E$

$$\mu_\beta(x) := \frac{1}{\mathcal{Z}_\beta} e^{-\beta V(x)} \lambda(x) \quad \text{with} \quad \mathcal{Z}_\beta := \sum_x e^{-\beta V(x)} \lambda(x) \quad (2.8)$$

The genealogical tree model associated with the genetic search model defined in (1.13), section 1.4, is defined by the line of ancestors of each individual  $\xi_n^i$  in the current population:

$$\forall 1 \leq i \leq N \quad \xi_{0,n}^i \longleftarrow \xi_{1,n}^i \longleftarrow \dots \longleftarrow \xi_{n-1,n}^i \longleftarrow \xi_{n,n}^i = \xi_n^i \quad (2.9)$$

We recall that this stochastic algorithm is defined in terms of a genetic type model with mutations transitions  $M_{\beta_n}$  and selection fitness functions  $G_n = e^{-(\beta_{n+1} - \beta_n)V}$ . For a more precise description of these ancestral lines we refer the reader to (1.4). In this situation, we can prove that  $\mathbb{Q}_n$  coincide with the limiting distribution of the ancestral lines of the genetic search algorithm, in the sense that

$$\forall 1 \leq i \leq N \quad \mathbb{Q}_n = \lim_{N \rightarrow \infty} \text{Law}(\xi_{0,n}^i, \xi_{1,n}^i, \dots, \xi_{n,n}^i) \quad (2.10)$$

In addition,  $\mathbb{Q}_n$  can also be interpreted as the limiting occupation measure of the genealogical tree, in the sense that:

$$\mathbb{Q}_n(f_n) = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N f_n(\xi_{0,n}^i, \xi_{1,n}^i, \dots, \xi_{n,n}^i) \quad (2.11)$$

Much more is true, we can prove that the ancestral lines are “almost” independent and identically distributed random paths with common distribution  $\mathbb{Q}_n$ . In much the same way, the measures  $\widehat{\mathbb{Q}}_n$  are approximated using the ancestral lines of the selected individuals  $\widehat{\xi}_n^i$  at time  $n$ .

In the end of section 2.5, we shall prove that the  $n$ -th marginal measures  $\eta_n$  defined in (2.6) also coincide with the Boltzmann-Gibbs measures (2.8) associated with an inverse temperature parameter  $\beta_n$ ; that is, we have that

$$\eta_n(x_n) = \mu_{\beta_n}(x) := \frac{1}{\mathcal{Z}_{\beta_n}} e^{-\beta_n V(x)} \lambda(x) \quad \text{and} \quad \mathcal{Z}_{\beta_n} = \prod_{0 \leq p < n} \eta_p(G_p) = \gamma_n(1) \quad (2.12)$$



(5) **Combinatorial counting and sampling**

Our next objective is to provide an interpretation of the measures  $\mathbb{Q}_n$  in terms of the genetic particle models presented in section 1.5. These models are defined as above by replacing in (2.8) the exponential Boltzmann-Gibbs measures by the measure  $\lambda$  restricted to a given non increasing sequence of subsets  $A_n \subset A_{n-1} \subset E$ ; that is, by the measures  $\mu_{A_n}$  defined below

$$\mu_{A_n}(x) := \frac{1}{\mathcal{Z}_{A_n}} 1_{A_n}(x) \lambda(x) \quad \text{with} \quad \mathcal{Z}_{A_n} := \sum_x 1_{A_n}(x) \lambda(x)$$

We recall that the corresponding stochastic algorithm is defined in terms of a genetic type model with mutations transitions  $M_{A_n}$  and selection potential functions  $G_n = 1_{A_{n+1}}$ , where  $M_{A_n}$  stands for a sequence of MCMC transitions with invariant measures  $\mu_{A_n}$ . The genealogical tree model associated with this genetic model is defined as in (2.9) and the pair of convergence results (2.10) and (2.11) are satisfied for the measures  $\mathbb{Q}_n$  associated with the Markov transitions  $M_{A_n}$  and the potential functions  $G_n = 1_{A_{n+1}}$ . Furthermore, the  $n$ -th time marginals measures  $\eta_n$  defined as in (2.6) are given by

$$\eta_n(f) := \gamma_n(f)/\gamma_n(1) \quad \text{with} \quad \gamma_n(f) = \mathbb{E} \left( f(X_n) \prod_{0 \leq p < n} 1_{A_{p+1}}(X_p) \right)$$

In the end of section 2.5, we shall also prove that

$$\eta_n(x_n) = \mu_{A_n}(x) := \frac{1}{\mathcal{Z}_{A_n}} 1_{A_n}(x) \lambda(x) \quad \text{and} \quad \mathcal{Z}_{A_n} = \prod_{0 \leq p < n} \eta_p(1_{A_{p+1}}) = \gamma_n(1) \tag{2.13}$$

(6) **Genetic search algorithms**

The interpretation of the measures  $\mathbb{Q}_n$  in terms of simple genetic models has already been developed in section 1.6. In this example, we recall that  $\mathbb{Q}_n$  can be seen as the asymptotic distribution of the ancestral lines of a simple genetic model with mutation transitions  $M_n$  and selection potential functions  $G_n$ . The  $n$ -th time marginal  $\eta_n$  of the measures  $\mathbb{Q}_n$  defined in (2.5) or in (2.6) can be regarded as the asymptotic distribution of an individual as the size of the population  $N$  tends to infinity. We can also interpret  $\eta_n$  as the limiting occupation measure of the genetic population at time  $n$ . These measures are often called the infinite population model in the literature on theoretical aspects of genetic models.

(7) **Directed polymers and self avoiding walks**

The interpretation of the measures  $\mathbb{Q}_n$  in terms of directed polymer models has already been developed in section 1.7. We simply notice that the probability measures  $\eta_n$  defined in (2.11) can be extended to path-valued Markov chains

$$X_n = (X'_0, \dots, X'_n) \in E_n := \underbrace{E \times \dots \times E}_{(n+1) \text{ times}}$$

where  $X'_n$  stands for an auxiliary  $E$ -valued Markov chain. In this situation, we observe that the potential functions  $G_n$  are defined on the path space  $E_n$ . Whenever  $X'_n$  is the simple random walk on the  $d$ -dimensional lattice  $E = \mathbb{Z}^d$ , for the choice of potential functions

$$\forall n \geq 0 \quad G_n(X_n) = G_n(X'_0, \dots, X'_n) = 1_{\mathbb{Z}^d - \{X'_0, \dots, X'_{n-1}\}}(X'_n)$$

we have that

$$\mathbb{Q}_n = \text{Law} \left( (X'_0, \dots, X'_n) \mid \forall 0 \leq p < q < n \quad X'_p \neq X'_q \right)$$

and

$$\widehat{\mathbb{Q}}_n = \text{Law} \left( (X'_0, \dots, X'_n) \mid \forall 0 \leq p < q \leq n \quad X'_p \neq X'_q \right)$$

In this situation, we recall that

$$\gamma_n(1) = \text{Proba} \left( \forall 0 \leq p < q < n \quad X'_p \neq X'_q \right) = \text{Card}(\mathbf{S}_{n-1}) / (2d)^{n-1}$$

where  $\mathbf{S}_{n-1}$  is the set of self-avoiding walks of length  $(n-1)$  starting at the origin.

## 2.5. Nonlinear recursive equations

This section is concerned with the analysis of the flow of probability measures  $\eta_n$  and  $\widehat{\eta}_n$  defined respectively in (2.6) and (2.7). First, using the Markov property of the signal, we readily prove that

$$\begin{aligned} \gamma_n(f) &= \mathbb{E} \left( \mathbb{E} \left( f(X_n) \prod_{0 \leq p < n} G_p(X_p) \mid X_0, \dots, X_{n-1} \right) \right) \\ &= \mathbb{E} \left( \mathbb{E}(f(X_n) \mid X_0, \dots, X_{n-1}) \prod_{0 \leq p < n} G_p(X_p) \right) \end{aligned}$$

from which we conclude that

$$\gamma_n(f) = \mathbb{E} \left( \mathbb{E}(f(X_n) \mid X_{n-1}) \prod_{0 \leq p < n} G_p(X_p) \right)$$

Recalling that  $\mathbb{E}(f(X_n) \mid X_{n-1}) = M_n(f)(X_{n-1})$ , we conclude that  $\gamma_n(f) = \widehat{\gamma}_{n-1} M_n(f)$ . In matrix notation, we conclude that

$$\gamma_n = \widehat{\gamma}_{n-1} M_n \tag{2.14}$$

On the other hand, we have

$$\widehat{\eta}_n(f) = \mathbb{E} \left( [f(X_n) G_n(X_n)] \prod_{0 \leq p < n} G_p(X_p) \right) = \gamma_n(f G_n) \tag{2.15}$$

with the multiplicative function  $(f G_n)(x_n) = f(x_n) G_n(x_n)$ . Our next objective is to derive the nonlinear recursions of the sequence of measures  $\eta_n$  and  $\widehat{\eta}_n$ . From previous formulae, we find that

$$\widehat{\eta}_n(f) = \frac{\gamma_n(G_n f)}{\gamma_n(G_n)} = \frac{\gamma_n(G_n f) / \gamma_n(1)}{\gamma_n(G_n) / \gamma_n(1)} = \frac{\eta_n(G_n f)}{\eta_n(G_n)}$$

This shows that  $\widehat{\eta}_n$  is deduced from  $\eta_n$  using the following formula

$$\widehat{\eta}_n := \Psi_{G_n}(\eta_n)$$

with the BBG transformation:

$$\Psi_{G_n}(\eta_n)(x_n) := \frac{1}{\eta_n(G_n)} G_n(x_n) \eta_n(x_n)$$

On the other hand, we have

$$\eta_n(f) = \frac{\widehat{\gamma}_{n-1} M_n(f)}{\widehat{\gamma}_{n-1}(1)} = \widehat{\eta}_{n-1} M_n(f)$$

from which we conclude that

$$\eta_n := \widehat{\eta}_{n-1} M_n$$

From these expressions, it is not difficult to establish the following proposition.

**Proposition 2.4.** *The measures  $\eta_n$  and  $\widehat{\eta}_n$  satisfy the following equations:*

$$\eta_n := \Psi_{G_{n-1}}(\eta_{n-1})M_n \quad \text{and} \quad \widehat{\eta}_n := \Psi_{G_n}(\widehat{\eta}_{n-1}M_n)$$

Another important observation is that the unnormalized distributions  $\gamma_n$  can be expressed in terms of the sequence of normalized measures from the origin up to the current time horizon. To be more precise, we have

$$\gamma_{n+1}(1) = \gamma_n(G_n) = \frac{\gamma_n(G_n)}{\gamma_n(1)} \gamma_n(1) = \eta_n(G_n) \gamma_n(1)$$

This multiplicative recursion yields the following integration exchange formula.

$$\gamma_{n+1}(1) = \mathbb{E} \left( \prod_{0 \leq p \leq n} G_p(X_p) \right) = \prod_{0 \leq p \leq n} \eta_p(G_p) \quad (2.16)$$

More generally, we have the following proposition which can be easily checked.

**Proposition 2.5.** *For any function  $f$ , and any time  $n \geq 1$ , we have*

$$\gamma_n(f) = \eta_n(f) \gamma_n(1) \quad \text{with} \quad \gamma_n(f) = \eta_n(f) \prod_{0 \leq p \leq n} \eta_p(G_p)$$

We end this section with a two-line proof of the pair of formulae (2.12) and (2.12). In the first situation, combining the fixed point equation with the BGG transformation given below

$$\eta_n = \eta_n M_{\beta_n} \quad \text{and} \quad \eta_n = \Psi_{G_{n-1}}(\eta_{n-1}) \quad \text{with} \quad G_{n-1} = \exp[-(\beta_n - \beta_{n-1})V]$$

we find that

$$\eta_n = \eta_n M_{\beta_n} = \Psi_{G_{n-1}}(\eta_{n-1}) M_{\beta_n}$$

Also observe that

$$\mathcal{Z}_{\beta_n} = \lambda(e^{-\beta_n V}) = \eta_{n-1}(G_{n-1}) \mathcal{Z}_{\beta_{n-1}} = \mathcal{Z}_0 \times \prod_{0 \leq p < n} \eta_p(G_p)$$

In much the same way, in the second situation, combining the fixed point equation with the BGG transformation given below

$$\eta_n = \eta_n M_{A_n} \quad \text{and} \quad \eta_n = \Psi_{G_{n-1}}(\eta_{n-1}) \quad \text{with} \quad G_{n-1} = 1_{A_n}$$

we find that

$$\eta_n = \eta_n M_{A_n} = \Psi_{G_{n-1}}(\eta_{n-1}) M_{A_n}$$

## 2.6. Nonlinear Markov chain models

In this section we design a nonlinear Markov interpretation of the flow of Feynman-Kac probability measures  $\eta_n$  associated with a pair of potential functions and Markov transitions  $(G_n, M_n)$ . To simplify the presentation, we further assume that the potential functions  $G_n$  take values in  $[0, 1]$ . More general models can be handled using the developments of section 2.2, dedicated to the transformations  $\Psi_{G_n}$  associated with more general potential functions  $G_n$ .

Combining proposition 2.4 with (2.2), we readily find that

$$\forall n \geq 0 \quad \eta_{n+1} = \Psi_{G_n}(\eta_n) M_{n+1} \quad \text{and} \quad \Psi_{G_n}(\eta_n) = \eta_n S_{n, \eta_n}$$

with the Markov transitions

$$S_{n,\eta_n}(x, y) = G_n(x) 1_x(y) + (1 - G_n(x)) \Psi_{G_n}(\eta_n)(y).$$

This implies that

$$\begin{aligned} \eta_{n+1}(z) &= \sum_y \left( \sum_x \eta_n(x) S_{n,\eta_n}(x, y) \right) M_{n+1}(y, z) \\ &= \sum_x \eta_n(x) \left( \sum_y S_{n,\eta_n}(x, y) M_{n+1}(y, z) \right) \end{aligned}$$

from which we conclude that

$$\eta_{n+1}(z) = \sum_x \eta_n(x) K_{n+1,\eta_n}(x, z)$$

with the Markov transitions  $K_{n+1,\eta_n}$  defined as the matrix composition of the selection type transition  $S_{n,\eta_n}$  with the free exploration Markov transition  $M_{n+1}$  given by

$$K_{n+1,\eta_n}(x, z) := \sum_y S_{n,\eta_n}(x, y) M_{n+1}(y, z)$$

We let  $(\bar{X}_n)_{n \geq 0}$  be a Markov chain on  $E$ , with initial distribution  $\eta_0 = \text{Law}(X_0)$  and elementary Markov transitions given by

$$\mathbb{P}(\bar{X}_{n+1} = y \mid \bar{X}_n = x) = K_{n+1,\eta_n}(x, dy) \quad \text{with} \quad \text{Law}(\bar{X}_n) = \eta_n$$

By a simple induction on the time parameter, we can easily check that  $\eta_n$  coincides with the distribution of the random states  $\bar{X}_n$ .

**Definition 2.6.** *The law of the paths of this non linear Markov chain*

$$\mathbb{P}((\bar{X}_0, \bar{X}_1, \dots, \bar{X}_n) = (x_0, x_1, \dots, x_n)) = \eta_0(x_0) K_{1,\eta_0}(x_0, x_1) \dots K_{n,\eta_{n-1}}(x_{n-1}, x_n)$$

is called the Mc-Kean measure associated with the Markov transitions  $K_{n+1,\eta_n}$ .

## 2.7. Feynman-Kac semigroup models

In section 2.1 we have seen that the distribution of a Markov chain on some finite space can be computed by simple matrix operations. In this case, if we denote the law of the chain  $\eta_p = \text{Law}(X_p)$  at time  $p$  then we obtain the law at any time  $n \geq p$  by an elementary composition of  $(n - p)$  transition matrices

$$\eta_n = \eta_p M_{p,n} \quad \text{with the matrix} \quad M_{p,n} := M_{p+1} M_{p+2} \dots M_n$$

For the flow of Feynman-Kac measures  $\eta_n$  defined in (2.6) the situation is more involved. By proposition 2.4, the measure  $\eta_n$  is connected to  $\eta_{n-1}$  by some transformation given by

$$\eta_n = \Phi_n(\eta_{n-1}) := \Psi_{G_{n-1}}(\eta_{n-1}) M_n$$

**Definition 2.7.** *The semigroup of the flow of Feynman-Kac measures  $\eta_n$  is given by the composition of the one step mappings  $\Phi_n$ ; that is, we have*

$$\forall 0 \leq p \leq n \quad \Phi_{p,n} = \Phi_n \circ \Phi_{n-1} \circ \dots \circ \Phi_{p+1}$$

with the convention  $\Phi_{n,n} = Id$  the identity mapping. By construction, for any pair of indexes  $0 \leq p \leq n$ , the measure  $\eta_p$  and  $\eta_n$  are connected with the following formula

$$\eta_n = \Phi_{p,n}(\eta_p)$$

Our next objective is to make more precise the definition given above.

First, using (2.14) and (2.15) we find that

$$\forall n \geq 0 \quad \gamma_n(y) = (\widehat{\gamma}_{n-1}M_n)(y) = \sum_x \widehat{\gamma}_{n-1}(x)M_n(x, y) \quad \text{and} \quad \widehat{\gamma}_n(x) = \gamma_n(x)G_n(x)$$

from which we conclude that

$$\gamma_n(y) = (\gamma_{n-1}Q_n)(y) = \sum_x \gamma_{n-1}(x)Q_n(x, y) \quad \text{with} \quad Q_n(x, y) = G_{n-1}(x)M_n(x, y)$$

In other words, we have proved that

$$\begin{aligned} \gamma_n &= \gamma_{n-1}Q_n = \gamma_{n-2}Q_{n-1}Q_n = \dots \\ &= \gamma_p Q_{p,n} \quad \text{with the matrix} \quad Q_{p,n} := Q_{p+1}Q_{p+2} \dots Q_n \end{aligned}$$

and therefore, for any function  $f$  on the set  $E$ , we have

$$\eta_n(f) := \frac{\gamma_n(f)}{\gamma_n(1)} = \frac{\gamma_p Q_{p,n}(f)}{\gamma_p Q_{p,n}(1)} = \frac{\gamma_p Q_{p,n}(f)/\gamma_p(1)}{\gamma_p Q_{p,n}(1)/\gamma_p(1)} = \frac{\eta_p Q_{p,n}(f)}{\eta_p Q_{p,n}(1)}$$

In summary, we have proved the following proposition.

**Proposition 2.8.** *For any function  $f$  on  $E$  we have*

$$\forall 0 \leq p \leq n \quad \Phi_{p,n}(\eta_p)(f) := \frac{\eta_p Q_{p,n}(f)}{\eta_p Q_{p,n}(1)}$$

### 3. INTERACTING PARTICLE ALGORITHMS

#### 3.1. Mean field particle interpretation models

In section 2.6 we provide a nonlinear Markov interpretation of the flow of Feynman-kac measures  $\eta_n$ . In this interpretation,  $\eta_n$  can be seen as the distribution of the random states  $\bar{X}_n$  of a Markov chain  $(\bar{X}_n)_{n \geq 0}$  on  $E$ , with initial distribution  $\eta_0 = \text{Law}(X_0)$  and elementary Markov transitions given by

$$\mathbb{P}(\bar{X}_{n+1} = y \mid \bar{X}_n = x) = K_{n+1, \eta_n}(x, dy) \quad \text{with} \quad \text{Law}(\bar{X}_n) = \eta_n$$

At a first level of analysis, let us suppose that the elementary transitions  $\bar{X}_{n-1} \rightsquigarrow \bar{X}_n$  are easy to sample. In this situation, we can approximate the flow of measures by sampling a series of  $N$  independent copies  $(\bar{X}_n^i)_{1 \leq i \leq N}$  of the chain  $\bar{X}_n$ . By the well-known law of large numbers, in some sense we have for any time horizon  $n$  and any function  $f$

$$\eta_n^N(f) := \frac{1}{N} \sum_{i=1}^N f(\bar{X}_n^i) \xrightarrow{N \uparrow \infty} \eta_n(f) = \mathbb{E}(f(\bar{X}_n))$$

In the above displayed formula we have used the notation

$$\forall x \in E \quad \eta_n^N(x) = \frac{1}{N} \sum_{i=1}^N 1_{\bar{X}_n^i}(x) \implies \eta_n^N(f) = \sum_x \eta_n^N(x) f(x) = \frac{1}{N} \sum_{i=1}^N f(\bar{X}_n^i)$$

More generally, for any function  $f_n$  on the path space  $E^{n+1}$ , we have

$$\frac{1}{N} \sum_{i=1}^N f_n(\bar{X}_0^i, \dots, \bar{X}_n^i) \xrightarrow{N \uparrow \infty} \mathbb{E}(f_n(\bar{X}_0, \dots, \bar{X}_n))$$

The schematic picture of this elementary Monte Carlo algorithm is given below

$$\begin{array}{ccc} \bar{X}_n^1 & \xrightarrow{K_{n+1, \eta_n}} & \bar{X}_{n+1}^1 \\ \vdots & & \vdots \\ \bar{X}_n^i & \xrightarrow{K_{n+1, \eta_n}} & \bar{X}_{n+1}^i \\ \vdots & & \vdots \\ \bar{X}_n^N & \xrightarrow{K_{n+1, \eta_n}} & \bar{X}_{n+1}^N \end{array}$$

Unfortunately, for most realistic models we can rarely compute explicitly the law  $\eta_n$  of random states of the chain  $\bar{X}_n$ . We can use the occupation measure of the  $N$  samples at every time step  $n$  to approximate the law  $\eta_n$ .

Under some appropriate regularity conditions on the Markov transitions  $K_{n+1, \eta_n(x, y)}$  we have

$$((\forall f : E \mapsto \mathbb{R}) \quad \eta_n^N(f) \simeq_{N \uparrow \infty} \eta_n(f)) \implies K_{n+1, \eta_n^N}(x, y) \simeq_{N \uparrow \infty} K_{n+1, \eta_n}(x, y)$$

We illustrate this rather abstract discussion with the Markov transitions  $K_{n+1, \eta_n}(x, z)$  discussed in section 2.6. We recall that in this case,  $K_{n+1, \eta_n}(x, z)$  is given by the following matrix composition

$$K_{n+1, \eta_n}(x, z) := \sum_y S_{n, \eta_n}(x, y) M_{n+1}(y, z) \quad (3.1)$$

with the Markov transitions

$$S_{n, \eta_n}(x, y) = G_n(x) 1_x(y) + (1 - G_n(x)) \Psi_{G_n}(\eta_n)(y).$$

and the BBG transformation

$$\Psi_{G_n}(\eta_n)(x) := \frac{1}{\eta_n(G_n)} G_n(x) \eta_n(x)$$

Replacing  $\eta_n$  by the occupation measure  $\eta_n^N(x) = \frac{1}{N} \sum_{i=1}^N 1_{\bar{X}_n^i}(x)$ , we find that

$$K_{n+1, \eta_n^N} := S_{n, \eta_n^N} M_{n+1} \quad \text{with} \quad S_{n, \eta_n^N}(x, y) = G_n(x) 1_x(y) + (1 - G_n(x)) \Psi_{G_n}(\eta_n^N)(y)$$

and the weighted BBG occupation measures

$$\Psi_{G_n}(\eta_n^N)(y) = \sum_{1 \leq j \leq N} \frac{G_n(\bar{X}_n^j)}{\sum_{1 \leq k \leq N} G_n(\bar{X}_n^k)} 1_{\bar{X}_n^j}(y)$$

We use this mean field approximation strategy inductively to define a Markov chain

$$(\xi_n^1, \xi_n^2, \dots, \xi_n^N)_{n \geq 0}$$

on the product state spaces  $E^N$ . The initial state  $(\xi_0^1, \xi_0^2, \dots, \xi_0^N)$  consists of  $N$  independent and identically distributed random variables with common distribution  $\eta_0$  termed particles. The elementary transitions are defined by the following synthetic diagram

$$\begin{array}{ccc} \xi_n^1 & \xrightarrow{K_{n+1, \eta_n^N}} & \xi_{n+1}^1 \\ \vdots & & \vdots \\ \xi_n^i & \xrightarrow{K_{n+1, \eta_n^N}} & \xi_{n+1}^i \\ \vdots & & \vdots \\ \xi_n^N & \xrightarrow{K_{n+1, \eta_n^N}} & \xi_{n+1}^N \end{array}$$

The mean field particle interpretation of this nonlinear measure valued model is an  $E^N$ -valued Markov chain

$$\xi_n = (\xi_n^1, \xi_n^2, \dots, \xi_n^N) \in E^N$$

with elementary transitions defined as

$$\mathbb{P}(\xi_{n+1} = (x^1, \dots, x^N) \mid \xi_n) = \prod_{i=1}^N K_{n+1, \eta_n^N}(\xi_n^i, x^i) \quad \text{with} \quad \eta_n^N := \frac{1}{N} \sum_{j=1}^N 1_{\xi_n^j} \quad (3.2)$$

In other words, given the population of  $N$  particles  $\xi_n = (\xi_n^1, \xi_n^2, \dots, \xi_n^N)$  at time  $n$ , the population of particles  $\xi_{n+1} = (\xi_{n+1}^1, \xi_{n+1}^2, \dots, \xi_{n+1}^N)$  at time  $(n+1)$  are independent random variables with respective distributions

$$K_{n+1, \eta_n^N}(\xi_n^1, x^1), \quad K_{n+1, \eta_n^N}(\xi_n^2, x^2), \quad \dots, \quad K_{n+1, \eta_n^N}(\xi_n^N, x^N)$$

The rationale behind this approximation is that  $\eta_{n+1}^N$  is the empirical measure associated with  $N$  independent variables with distributions  $K_{n+1, \eta_n^N}(\xi_n^i, x)$ , so as soon as  $\eta_n^N$  is a good approximation of  $\eta_n$  then, in view of (3.2),  $\eta_{n+1}^N$  should be a good approximation of  $\eta_{n+1}$ .

In the situation where  $K_{n+1, \eta_n}$  is given by the (3.1), the flow of Feynman-Kac measures evolves according to the two-step updating/prediction transitions,

$$\eta_n \xrightarrow{S_{n, \eta_n}} \hat{\eta}_n = \eta_n S_{n, \eta_n} = \Psi_{G_n}(\eta_n) \xrightarrow{M_{n+1}} \eta_{n+1} = \hat{\eta}_n M_{n+1} \quad (3.3)$$

In the  $N$ -mean field particle interpretation model, this pair of recursions has been replaced by a two-step selection/mutation transition in product spaces

$$\xi_n \in E^N \xrightarrow{\text{selection}} \hat{\xi}_n \in E^N \xrightarrow{\text{mutation}} \xi_{n+1} \in E^N \quad (3.4)$$

The genetic type evolution of the system is summarized by the following synthetic picture:

$$\left[ \begin{array}{c} \xi_n^1 \\ \vdots \\ \xi_n^i \\ \vdots \\ \xi_n^N \end{array} \right] \xrightarrow{S_{n, \eta_n^N}} \left[ \begin{array}{ccc} \hat{\xi}_n^1 & \xrightarrow{M_{n+1}} & \xi_{n+1}^1 \\ \vdots & & \vdots \\ \hat{\xi}_n^i & \longrightarrow & \xi_{n+1}^i \\ \vdots & & \vdots \\ \hat{\xi}_n^N & \longrightarrow & \xi_{n+1}^N \end{array} \right]$$

with the selection Markov transition :

$$S_{n,\eta_n^N}(\xi_n^i, x) := G_n(\xi_n^i) 1_{\xi_n^i}(x) + (1 - G_n(\xi_n^i)) \sum_{1 \leq j \leq N} \frac{G_n(\xi_n^j)}{\sum_{1 \leq k \leq N} G_n(\xi_n^k)} 1_{\xi_n^j}(x)$$

Using the Feynman-Kac modeling techniques presented in section 2.4, it is now easy to check that the corresponding mean field particle interpretations discussed above coincide with the algorithms discussed in section 1.

### 3.2. Some particle estimates

In section 3.1 we have designed a mean field particle interpretation of the flow of Feynman-Kac measures

$$\eta_n(f) := \gamma_n(f)/\gamma_n(1) \quad \text{with} \quad \gamma_n(f) = \mathbb{E} \left( f(X_n) \prod_{0 \leq p < n} G_p(X_p) \right)$$

We also recall from (2.16) that the normalizing constants  $\gamma_n(1)$  are expressed in terms of the flow of measures  $(\eta_p)_{0 \leq p < n}$  with the following multiplicative formula

$$\gamma_n(1) = \mathbb{E} \left( \prod_{0 \leq p < n} G_p(X_p) \right) = \prod_{0 \leq p < n} \eta_p(G_p)$$

This also implies that the unnormalized Feynman-Kac measures can be rewritten in the following form:

$$\gamma_n(f) = \eta_n(f) \gamma_n(1) \quad \text{with} \quad \gamma_n(f) = \eta_n(f) \prod_{0 \leq p < n} \eta_p(G_p)$$

The mean field particle interpretation of these measures presented in section 3.1 can be seen as genetic type algorithms with  $M_n$ -mutation transitions and  $G_n$ -potential functions, where  $M_n$  represents the Markov transition of the reference Markov chain  $X_n$ . Inversely, any genetic type model of that forms converge, in some sense, as the size of the population  $N$  tends to infinity to the above Feynman-Kac measures.

More formally, for any function  $f_n$  on the product space  $E^{(n+1)}$ , we have that

$$\begin{aligned} & \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N f_n(\xi_0^i, \xi_1^i, \dots, \xi_n^i) \\ &= \sum_{x_0, \dots, x_n} \eta_0(x_0) K_{1,\eta_0}(x_0, x_1) \dots K_{n,\eta_{n-1}}(x_{n-1}, x_n) f_n(x_0, \dots, x_n) \end{aligned}$$

and for any function  $f$  on  $E$

$$\lim_{N \rightarrow \infty} \eta_n^N(f) = \eta_n(f) \quad \text{with} \quad \eta_n^N(f) = \frac{1}{N} \sum_{i=1}^N f_n(\xi_n^i) \quad (3.5)$$

The normalizing constants  $\gamma_n(1) = \lim_{N \rightarrow \infty} \gamma_n^N(1)$  can be computed using the unbiased particle estimates  $\gamma_n^N(1)$  defined below

$$\gamma_n^N(1) := \prod_{0 \leq p < n} \eta_p^N(G_p) \quad \text{with} \quad \forall n \geq 0 \quad \eta_n^N(G_n) = \frac{1}{N} G_n(\xi_n^i)$$

To give a flavor of some non asymptotic variance estimates that can be obtained for such particle estimates, we present a recent result for time homogeneous models  $(G_n, M_n) = (G, M)$ .



**Theorem 3.1** ([5]). *Suppose that the pair of potential-transitions  $(G, M)$  are chosen so that*

$$\forall (x, x', y) \in E^3 \quad G(x) \leq \delta G(x') \quad \text{and} \quad M^m(x, y) \leq \beta M^m(x', y) \quad (3.6)$$

for some  $m \geq 1$  and some parameters  $(\delta, \beta) \in [1, \infty]^2$ . In this situation, any  $n \geq 0$ , and any  $N > (n+1)\beta\delta^m$  we have

$$\mathbb{E} \left[ (\gamma_n^N(1) - \gamma_n(1))^2 \right] \leq \gamma_n(1)^2 \left( \frac{4}{N} (n+1) \beta \delta^m \right) \quad (3.7)$$

More general results can be found in [9, 10], under weaker regularity conditions. In this connection, we mention that the  $N$ -particle approximation of the unnormalized measures  $\gamma_n$  is given by the unbiased estimates

$$\gamma_n^N(f) := \eta_n^N(f) \times \gamma_n^N(1) \xrightarrow{N \rightarrow \infty} \gamma_n(f) = \eta_n(f) \times \gamma_n(1)$$

At the end of this section, we discuss some path space particle estimates associated to these genetic algorithms. First, it is important to observe that the above constructions can be extended to Feynman-Kac models with reference Markov chains and potential functions on path spaces

$$X_n := (X'_0, \dots, X'_n) \in E_n = (E')^{(n+1)} \quad \text{and} \quad G_n(X_n) := G_n(X'_0, \dots, X'_n) \quad (3.8)$$

This observation is essential for the analysis of the directed polymers and the self avoiding walks discussed in section 1.7. In this situation, every particle at time  $n$  is a path-valued particle of length  $n$ . The selection transition consists of selecting the path-particle with high potential value, while the mutation transition simply consists of a path extension with an elementary move according to the auxiliary process  $X'_n$ , with Markov transitions  $M'_n$  on the state space  $E'$ .

When the potential functions only depend on the terminal value of the paths

$$G_n(X_n) := G'_n(X'_n)$$

for some fitness function  $G'_n$ , we can check that the path-valued particle model represents the time evolution of the genealogical tree model associated with the time evolution of the individuals  $\xi_n^i$  evolving with  $M'_n$ -mutations and  $G'_n$ -selections. In this situation, if

$$\xi_n^i := (\xi_{0,n}^i, \xi_{1,n}^i, \dots, \xi_{n,n}^i)$$

stands for the  $i$ -th ancestral line of the current individual  $\xi_{n,n}^i$  after the  $n$ -th mutation, then for any function  $f_n$  on  $E_n$ , we have that

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N f_n(\xi_{0,n}^i, \xi_{1,n}^i, \dots, \xi_{n,n}^i) = \frac{\mathbb{E} \left( f_n(X'_0, \dots, X'_n) \prod_{0 \leq p < n} G'_p(X'_p) \right)}{\mathbb{E} \left( \prod_{0 \leq p < n} G'_p(X'_p) \right)} \quad (3.9)$$

Note that the  $n$ -th time marginals converge to the Feynman-Kac measures associated to the pair  $(G'_n, M'_n)$ ; that is, for any function  $f$  on the state space  $E'$  we have

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N f(\xi_n^i) = \eta'_n(f) = \frac{\gamma'_n(f)}{\gamma'_n(1)} \quad \text{with} \quad \gamma'_n(f) := \mathbb{E} \left( f(X'_n) \prod_{0 \leq p < n} G'_p(X'_p) \right) \quad (3.10)$$

Another quantity of interest is the complete ancestral tree defined by the whole population of individuals from the origin up to the current time horizon

$$\forall 1 \leq i \leq N \quad (\xi_{0,0}^i, \xi_{1,1}^i, \dots, \xi_{n-1,n-1}^i, \xi_{n,n}^i)$$

The occupation measures of this complete ancestral converge to the Mc-Kean distributions of the nonlinear Markov chain with initial distribution  $\eta_0 = \eta'_0$  and elementary transitions  $K'_{n+1, \eta'_n}(x, y)$  on  $E'$  defined as the transitions  $K_{n+1, \eta_n}$  by replacing the triplet  $(\eta_n, G_n, M_n)$  by the triplet  $(\eta'_n, G'_n, M'_n)$ . In other words, we have

$$\begin{aligned} & \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N f_n(\xi_{0,0}^i, \xi_{1,1}^i, \dots, \xi_{n,n}^i) \\ &= \sum_{x_0, \dots, x_n} \eta'_0(x_0) K'_{1, \eta'_0}(x_0, x_1) \dots K'_{n, \eta'_{n-1}}(x_{n-1}, x_n) f_n(x_0, \dots, x_n) \end{aligned}$$

### 3.3. An elementary proof

In this section, we provide a simple proof of the convergence result (3.5) using a rather elementary induction w.r.t. the time parameter  $n$ . To simplify the analysis we assume that the potential functions  $G_n$  are chosen so that  $\delta(G_n) := \sup_{x,y} (G_n(x)/G_n(y)) < \infty$  and  $E$  is a finite set. The main result of this section is the following elementary variance estimates.

**Proposition 3.2.** *For any function  $f$  on  $E$ , any time horizon  $n \geq 0$ , and any population size  $N$ , we have the variance estimate*

$$\mathbb{E} \left( [\eta_n^N(f) - \eta_n(f)]^2 \right) \leq \frac{c(n)}{N} \text{osc}(f)^2 \quad (3.11)$$

for some finite constant  $c(n) < \infty$ , whose values do not depend on the parameter  $N$ .

#### 3.3.1. The initial particle approximation

We observe that the initial population  $\xi_0 = (\xi_0^1, \xi_0^2, \dots, \xi_0^N)$  consists of  $N$  independent and identically distributed random variables with common distribution  $\eta_0$ . For any function  $f$  on  $E$ , we have

$$\eta_0^N(f) = \eta_0(f) + \frac{1}{\sqrt{N}} V_0^N(f)$$

with the random quantities

$$V_0^N(f) := \frac{1}{\sqrt{N}} \sum_{1 \leq i \leq N} [f(\xi_0^i) - \eta_0(f)] = \sum_x V_0^N(x) f(x)$$

and the random occupation measures

$$V_0^N(x) = \frac{1}{\sqrt{N}} \sum_{1 \leq i \leq N} [1_{\xi_0^i}(x) - \eta_0(x)]$$

Using the decomposition

$$V_0^N(f)^2 = \frac{1}{N} \sum_{1 \leq i \leq N} [f(\xi_0^i) - \eta_0(f)]^2 + \frac{1}{N} \sum_{1 \leq i \neq j \leq N} [f(\xi_0^i) - \eta_0(f)] [f(\xi_0^j) - \eta_0(f)]$$

it is now a simple exercise to check that

$$\mathbb{E} \left( [V_0^N(f)]^2 \right) = \eta_0([f - \eta_0(f)]^2) = \eta_0(f^2) - \eta_0(f)^2 \leq \text{osc}(f)^2$$

where  $\text{osc}(f)$  stands for the oscillation of the function  $f$

$$\text{osc}(f) = \sup_{x,y} |f(x) - f(y)|$$

This shows that the initial distribution  $\eta_0$  is approximated using the initial occupation measure  $\eta_0^N$ , and we have the variance estimates

$$\mathbb{E} \left( [\eta_0^N(f) - \eta_0(f)]^2 \right) = \frac{1}{N} \eta_0([f - \eta_0(f)]^2) \leq \frac{1}{N} \text{osc}(f)$$

### 3.3.2. The local sampling errors

We recall that the population of particle at time  $(n+1)$  is the sequence of conditionally independent random variables

$$\xi_{n+1} = (\xi_{n+1}^1, \xi_{n+1}^2, \dots, \xi_{n+1}^N)$$

with respective distributions

$$K_{n+1, \eta_n^N}(\xi_n^1, x^1), \quad K_{n+1, \eta_n^N}(\xi_n^2, x^2), \quad \dots \quad K_{n+1, \eta_n^N}(\xi_n^N, x^N)$$

We observe that for any  $y \in E$  and any  $n \geq 0$ , we have

$$\mathbb{E}(\eta_{n+1}^N(f) | \xi_n) = \frac{1}{N} \sum_{1 \leq i \leq N} \mathbb{E}(f(\xi_{n+1}^i) | \xi_n) = \frac{1}{N} \sum_{1 \leq i \leq N} K_{n+1, \eta_n^N}(f)(\xi_n^i)$$

with the function

$$K_{n+1, \eta_n^N}(f)(x) = \sum_y K_{n+1, \eta_n^N}(x, y) f(y)$$

This implies that

$$\mathbb{E}(\eta_{n+1}^N(f) | \xi_n) = \eta_n^N K_{n+1, \eta_n^N}(f) = \Phi_{n+1}(\eta_n^N)(f)$$

with the probability measures

$$\Phi_{n+1}(\eta_n^N)(y) = (\eta_n^N K_{n+1, \eta_n^N})(y) = \sum_x \eta_n^N(x) \times K_{n+1, \eta_n^N}(x, y)$$

On the other hand, for any function  $f$  on  $E$ , we have the decomposition

$$\begin{aligned} \eta_{n+1}^N(f) - \Phi_{n+1}(\eta_n^N)(f) &= \eta_{n+1}^N(f) - \eta_n^N K_{n+1, \eta_n^N}(f) \\ &= \frac{1}{N} \sum_{1 \leq i \leq N} [f(\xi_{n+1}^i) - K_{n+1, \eta_n^N}(f)(\xi_n^i)] = \frac{1}{\sqrt{N}} V_{n+1}^N(f) \end{aligned}$$

In the above displayed formulae,  $V_{n+1}^N(f)$  stands for the random variable defined below

$$V_{n+1}^N(f) := \frac{1}{\sqrt{N}} \sum_{1 \leq i \leq N} [f(\xi_{n+1}^i) - K_{n+1, \eta_n^N}(f)(\xi_n^i)] = \sum_x V_{n+1}^N(x) f(x)$$

with the random measures

$$V_{n+1}^N(x) = \frac{1}{\sqrt{N}} \sum_{1 \leq i \leq N} [1_{\xi_{n+1}^i}(x) - K_{n+1, \eta_n^N}(\xi_n^i, x)]$$

Rewritten in a different way, we have

$$\eta_{n+1}^N = \Phi_{n+1}(\eta_n^N) + \frac{1}{\sqrt{N}} V_{n+1}^N$$

while the limiting measures satisfy the following equation

$$\eta_{n+1} = \Phi_{n+1}(\eta_n)$$

These decompositions show that the random measures  $V_n^N$  can be interpreted as the local sampling errors induced by the mean field particle transitions. The above formulae also show that the  $N$ -particle model can also be interpreted as a stochastic perturbation of the limiting system.

Using the fact that

$$\begin{aligned} [V_{n+1}^N(f)]^2 &= \frac{1}{N} \sum_{1 \leq i \leq N} [f(\xi_{n+1}^i) - K_{n+1, \eta_n^N}(f)(\xi_n^i)]^2 \\ &\quad + \frac{1}{N} \sum_{1 \leq i \neq j \leq N} [f(\xi_{n+1}^i) - K_{n+1, \eta_n^N}(f)(\xi_n^i)] \times [f(\xi_{n+1}^j) - K_{n+1, \eta_n^N}(f)(\xi_n^j)] \end{aligned}$$

It is rather elementary to check that

$$\mathbb{E}(V_{n+1}^N(f) \mid \xi_n) = 0 \tag{3.12}$$

and

$$\mathbb{E}\left([V_n^N(f)]^2 \mid \xi_n\right) = \frac{1}{N} \sum_{1 \leq i \leq N} \mathbb{E}\left([f(\xi_{n+1}^i) - K_{n+1, \eta_n^N}(f)(\xi_n^i)]^2 \mid \xi_n\right) \leq \text{osc}(f)^2$$

From these calculations, we readily prove the following local error estimates.

**Proposition 3.3.** *For any function  $f$  on  $E$ , any time horizon  $n \geq 0$ , and any population size  $N$ , we have the almost sure estimates*

$$\mathbb{E}\left([\eta_{n+1}^N(f) - \Phi_{n+1}(\eta_n^N)(f)]^2 \mid \xi_n\right)^{\frac{1}{2}} \leq \frac{1}{\sqrt{N}} \text{osc}(f)$$

### 3.3.3. The proof by induction

The main object of this section is to complete the proof of proposition 3.2.

By definition of the one step mappings, for any function  $f$  on  $E$ , any time horizon  $n \geq 0$ , and any population size  $N$ , we have

$$\Phi_{n+1}(\eta_n^N)(f) = \eta_n^N(G_n M_{n+1}(f)) / \eta_n^N(G_n)$$

and therefore

$$\Phi_{n+1}(\eta_n^N)(f) - \Phi_{n+1}(\eta_n)(f) = \eta_n^N(\tilde{f}_n) / \eta_n^N(\tilde{g}_n)$$

with the functions

$$\tilde{f}_n(x) := \tilde{g}_n(x) \times M_{n+1}[f - \Phi_{n+1}(\eta_n)(f)](x) \quad \text{and} \quad \tilde{g}_n(x) := \frac{G_n(x)}{\eta_n(G_n)} \tag{3.13}$$

Notice that the function  $\tilde{f}_n$  is such that

$$\eta_n(\tilde{f}_n) = 0 \quad \eta_n(\tilde{g}_n) = 1 \quad \text{and} \quad \|\tilde{f}_n\| := \sup_x |\tilde{f}_n(x)| \leq \delta(G_n) \text{osc}(f)$$

From the above observations, we prove the following decomposition

$$\Phi_{n+1}(\eta_n^N)(f) - \Phi_{n+1}(\eta_n)(f) = \frac{1}{\eta_n^N(\tilde{g}_n)} [\eta_n^N(\tilde{f}_n) - \eta_n(\tilde{f}_n)] \tag{3.14}$$

This implies that

$$|\Phi_{n+1}(\eta_n^N)(f) - \Phi_{n+1}(\eta_n)(f)| \leq \delta(G_n) \times |\eta_n^N(\tilde{f}_n) - \eta_n(\tilde{f}_n)|$$

and therefore

$$\mathbb{E} \left( [\Phi_{n+1}(\eta_n^N)(f) - \Phi_{n+1}(\eta_n)(f)]^2 \right) \leq \delta(G_n)^2 \mathbb{E} \left( [\eta_n^N(\tilde{f}_n) - \eta_n(\tilde{f}_n)]^2 \right)$$

We end the proof of proposition 3.2 using the following decomposition

$$\eta_{n+1}^N(f) - \eta_{n+1}(f) = [\eta_{n+1}^N(f) - \Phi_{n+1}(\eta_n^N)(f)] + [\Phi_{n+1}(\eta_n^N)(f) - \Phi_{n+1}(\eta_n)(f)]$$

Using the conditional unbiased property (3.12) and the local estimates stated in proposition 3.3, we prove that

$$\begin{aligned} & \mathbb{E} \left( [\eta_{n+1}^N(f) - \eta_{n+1}(f)]^2 \right) \\ &= \mathbb{E} \left( [\eta_{n+1}^N(f) - \Phi_{n+1}(\eta_n^N)(f)]^2 \right) + \mathbb{E} \left( [\Phi_{n+1}(\eta_n^N)(f) - \Phi_{n+1}(\eta_n)(f)]^2 \right) \\ &\leq \frac{1}{N} \text{osc}(f)^2 + \delta(G_n)^2 \mathbb{E} \left( [\eta_n^N(\tilde{f}_n) - \eta_n(\tilde{f}_n)]^2 \right) \end{aligned}$$

The end of the proof of the variance estimates stated in proposition 3.2 is now clear. In section 3.3.1 we have seen that (3.11) is true at rank  $n = 0$ . Assuming that (3.11) is true at rank  $n$ , the above inequality implies that the result is also satisfied at rank  $(n + 1)$ . This ends the proof of the variance estimates stated in the proposition.

#### 3.4. $\mathbb{L}_q$ -mean error and bias estimates

We extend in this section the variance estimates stated in proposition 3.2 to any  $\mathbb{L}_q$ -mean errors bounds. The forthcoming analysis is developed under the regularity conditions stated in the beginning of section 3.3.

**Proposition 3.4.** *For any function  $f$  on  $E$  with  $\text{osc}(f) \leq 1$ , any time horizon  $n \geq 0$ , any population size  $N$ , we have*

$$|\mathbb{E}(\eta_n^N(f)) - \eta_n(f)| \leq c(n)/N \quad (3.15)$$

for some finite constant  $c(n) < \infty$ . In addition, for any integer  $q \geq 1$ , we have the variance estimate

$$\sqrt{N} \mathbb{E} \left( |\eta_n^N(f) - \eta_n(f)|^q \right)^{\frac{1}{q}} \leq b(q) c(n)$$

for some finite constant  $c(n) < \infty$ , and the collection of constants  $b(q)$  defined below for any  $p \geq 0$ :

$$\begin{aligned} b(2p)^{2p} &:= (2p)_p 2^{-p} \\ b(2p+1)^{2p+1} &:= \frac{(2p+1)_{(p+1)}}{\sqrt{p+1/2}} 2^{-(p+1/2)} \quad \text{with} \quad (q+p)_p := (q+p)!/p! \end{aligned} \quad (3.16)$$

In the further development of this section, we provide a rather elementary proof of the above proposition based on a simple induction on the time parameter combined with Kintchine's type inequalities. The resulting estimates of the upper bound constants  $c(n)$  are far from being "sharp". Next, we present without proofs some more precise extensions of these results. Detailed proofs of these results can be found in [9, 10].

- (1) When the flow of measures is stable, in the sense that it forgets exponentially fast its initial conditions, we can prove uniform estimates w.r.t. the time horizon. For instance, under the regularity condition (3.6) for some parameters  $(m, \delta, \beta)$ , we can prove that the  $\mathbb{L}_q$ -mean error estimates (3.11) holds with some constants  $c(n)$  such that

$$\sup_{n \geq 0} c(n) \leq 2m\delta^{2m-1}\beta^3$$

- (2) Similar  $\mathbb{L}_q$ -mean error estimates (3.11) can be derived for the convergence of the occupation measures (3.9) associated with genealogical tree models. Up to a state space enlargement the convergence analysis of models associated with path-valued Markov chains (3.8) follows exactly the same line of arguments. In addition, when the flow of marginal measures  $\eta'_n$  introduced in (3.10) is stable, we can prove  $\mathbb{L}_q$ -mean errors bounds (3.11) for the occupation measures of genealogical tree models with some constant that grows linearly w.r.t. the time parameter; that is, such that  $c(n) = c n$ , form some finite constant  $c < \infty$ .
- (3) The extensions of the  $\mathbb{L}_q$ -mean errors bounds (3.11) to general state space models and Zolotarev's type seminorms

$$\|\eta_n^N - \eta_n\|_{\mathcal{F}} := \sup_{f \in \mathcal{F}} |\eta_n^N(f) - \eta_n(f)|$$

associated with some collection  $\mathcal{F}$  of functions can also be found in [9].

- (4) Note that the mean value of the occupation measures coincides with the law of a given particle; that is, we have that

$$\mathbb{E}(\eta_n^N(f)) = \frac{1}{N} \sum_{1 \leq i \leq N} \mathbb{E}(f(\xi_n^i)) = \mathbb{E}(f(\xi_n^1))$$

For finite state space models, the bias estimates (3.15) states that

$$\|\text{Law}(\xi_n^1) - \eta_n\|_{\text{tv}} := \sum_x |\mathbb{P}(\xi_n^1 = x) - \eta_n(x)| \leq c(n)/N$$

One again, the constant  $c(n)$  can be uniformly bounded w.r.t. the time parameter under appropriate regularity conditions. Furthermore, similar estimates holds true for the ancestral line of the corresponding genealogical tree models. Notice that in this situation  $\eta_n$  is replaced by the Mc-Kean measure introduced in definition 2.6.

All of these estimates can be extended to general state space models and to the law of the first  $q$  particles in the population

$$\left\| \text{Law}(\xi_n^1, \dots, \xi_n^q) - \text{Law}(\bar{X}_n^1, \dots, \bar{X}_n^q) \right\|_{\text{tv}} \leq c(n) q^2/N$$

where  $\bar{X}_n^i$  is a sequence of independent and identically distributed random variables with common distribution  $\eta_n$ . This property is called the propagation of chaos property in applied probability.

We end this section with a pair of Bernstein's type exponential concentration estimates recently obtained in [13]. For any  $N \geq 1$ , any  $n \geq 0$ , and any  $\lambda > 0$ , we have

$$-\frac{1}{N} \log \mathbb{P} \left( |\eta_n^N - \eta_n|(f) \geq \frac{r}{N} + \lambda \right) \geq \frac{\lambda^2}{2} \left( \left( \bar{\sigma} + \frac{\sqrt{2}r}{\sqrt{N}} \right)^2 + 2\lambda \left( r + \frac{\beta\delta^m}{3} \right) \right)^{-1}$$

and

$$-\frac{1}{N} \log \mathbb{P} \left( |\eta_n^N - \eta_n|(f) \geq \frac{r}{N} + \lambda \right) \geq \frac{\lambda^2}{2} \left( \left( \hat{\sigma} + \frac{\sqrt{2}r}{\sqrt{N}} \right)^2 + 2\lambda r \right)^{-1}.$$

with some parameters  $(r, \bar{\sigma}, \hat{\sigma})$  such that

$$r \leq 4m\delta^{4m-1}\beta^5 \quad \bar{\sigma}^2 \leq 16m\delta^{5m-1}\beta^6 \sigma^2 \quad \text{and} \quad \hat{\sigma}^2 \leq 4m\delta^{3m-1}\beta^4$$

with the uniform local variance parameters:

$$\sigma^2 := \sup \mu \left( K_{n,\mu} [f_n - K_{n,\mu}(f_n)]^2 \right) (\leq 1).$$

In the above displayed formula the supremum is taken over all integers  $n \geq 1$ , any probability measures and all functions  $f$  on  $E$  with  $\text{osc}(f) \leq 1$ .

To see how these parameters enter into the local fluctuation variance, we notice that for any probability measure we have

$$\mu [K_{n,\mu} ((f - K_{n,\mu}(f))^2)] = \Phi_n(\mu) [(f - \Phi_n(\mu)(f))^2] - \mu \left[ [G_{n-1}(Id - \Psi_n(\mu))(M_n(f))]^2 \right]$$

We also have the following formula

$$\mu [K_{n,\mu} ((f - K_{n,\mu}(f))^2)] = \mu \Gamma_{L_{n,\mu}}(f, f) - \mu [L_{n,\mu}(f)^2]$$

in terms of the operators

$$\begin{aligned} L_{n,\mu}(f) &= G_{n-1} (M_n - Id)(f) \\ &\quad + (1 - G_{n-1}) [(\Psi_{n-1}(\mu) - Id) + \Psi_{n-1}(\mu)(M_n - Id)](f) \end{aligned}$$

and their ‘‘carré du champ’’ function  $\Gamma_{L_{n,\mu}}(f, f)$  defined for any  $x \in E$  by

$$\Gamma_{L_{n,\mu}}(f, f)(x) = L_{n,\mu} \left( [f - L_{n,\mu}(f)(x)]^2 \right) (x) = L_{n,\mu}(f^2)(x) - 2f(x) L_{n,\mu}(f)(x)$$

#### 3.4.1. Kintchine’s type inequalities

The  $\mathbb{L}_m$ -mean errors analysis developed in this section is based on the following Kintchine type inequality for martingales with symmetric and independent increments.

**Lemma 3.5.** *Let  $M_n := \sum_{0 \leq p \leq n} \Delta_p$  be a real valued martingale with symmetric and independent increments  $(\Delta_n)_{n \geq 0}$ . For any integer  $m \geq 1$ , and any  $n \geq 0$ , we have*

$$\mathbb{E}(|M_n|^m)^{\frac{1}{m}} \leq b(m) \mathbb{E} \left( [M_n]_n^{m'/2} \right)^{\frac{1}{m'}} \quad (3.17)$$

with the smallest even integer  $m' \geq m$ , the bracket process  $[M]_n := \sum_{0 \leq p \leq n} \Delta_p^2$ , and the collection of constants  $b(m)$  defined in (3.16).

#### Proof:

We prove the lemma by induction on the parameter  $n$ . The result is clearly satisfied for  $n = 0$ . Suppose the estimate (3.17) is true at rank  $(n - 1)$ . To prove the result at rank  $n$ , we use the binomial decomposition

$$(M_{n-1} + \Delta_n)^{2m} = \sum_{p=0}^{2m} \binom{2m}{p} M_{n-1}^{2m-p} (\Delta_n)^p$$

Using the symmetry condition, all the odd moments of  $\Delta_n$  are null. Consequently, we find that

$$\mathbb{E} \left( (M_{n-1} + \Delta_n)^{2m} \right) = \sum_{p=0}^m \binom{2m}{2p} \mathbb{E} \left( M_{n-1}^{2(m-p)} \right) \mathbb{E} (\Delta_n^{2p})$$

Using the induction hypothesis, we prove that the above expression is upper bounded by the quantity

$$\sum_{p=0}^m \binom{2m}{2p} 2^{-(m-p)} (2(m-p))_{(m-p)} \mathbb{E} \left( [M]_{n-1}^{m-p} \right) \mathbb{E} (\Delta_n^{2p})$$

To take the final step, we use the fact that

$$\binom{2m}{2p} 2^{-(m-p)} (2(m-p))_{(m-p)} = \frac{2^{-m} (2m)_m}{2^{-p} (2p)_p} \binom{m}{p} \quad \text{and} \quad (2p)_p \geq 2^p$$

to conclude that

$$\begin{aligned} \mathbb{E} \left( (M_{n-1} + \Delta_n)^{2m} \right) &\leq 2^{-m} (2m)_m \sum_{p=0}^m \binom{m}{p} \mathbb{E} \left( [M]_{n-1}^{m-p} \right) \mathbb{E} \left( \Delta_n^{2p} \right) \\ &= 2^{-m} (2m)_m \mathbb{E} \left( [M]_n^m \right) \end{aligned}$$

For odd integers we use twice the Cauchy-Schwarz inequality to deduce that

$$\begin{aligned} \mathbb{E}(|M_n|^{2m+1})^2 &\leq \mathbb{E}(M_n^{2m}) \mathbb{E}(M_n^{2(m+1)}) \\ &\leq 2^{-(2m+1)} (2m)_m (2(m+1))_{(m+1)} \mathbb{E} \left( [M]_n^{m+1} \right)^{\frac{2m+1}{m+1}} \end{aligned}$$

Since we also have

$$(2(m+1))_{(m+1)} = \frac{(2(m+1))!}{(m+1)!} = 2 \frac{(2m+1)!}{m!} = 2 (2m+1)_{(m+1)}$$

and

$$(2m)_m = \frac{1}{2m+1} \frac{(2m+1)!}{m!} = \frac{1}{2m+1} (2m+1)_{(m+1)}$$

we conclude that

$$\mathbb{E}(|M_n|^{2m+1}) \leq 2^{-(m+1/2)} \frac{(2m+1)_{(m+1)}}{\sqrt{m+1/2}} \mathbb{E} \left( [M]_n^{m+1} \right)^{1 - \frac{1}{2(m+1)}}$$

This ends the proof of the lemma. ■

### 3.4.2. Initial and local sampling estimates

By construction, for any function  $f$  on the set  $E$  and any  $n \geq 0$  we have

$$V_n^N(f) = \sum_{i=1}^N \Delta_{n,i}^{(N)}(f) \quad \text{with} \quad \Delta_{n,i}^{(N)}(f) := \frac{1}{\sqrt{N}} \left[ f(\xi_n^i) - K_{n,\eta_{n-1}^N}(f)(\xi_{n-1}^i) \right]$$

For  $n = 0$ , we use the convention  $K_{n,\eta_{n-1}^N}(f)(\xi_{n-1}^i) = \eta_0(f)$ , so that

$$V_0^N(f) = \sqrt{N} \left[ \eta_0^N(f) - \eta_0(f) \right]$$

Given  $\xi_{n-1}$ , we let  $(\zeta_n^i)_{1 \leq i \leq N}$  be an independent copy of  $(\xi_n^i)_{1 \leq i \leq N}$ . We readily check that

$$\Delta_{n,i}^{(N)}(f) := \mathbb{E} \left( \frac{1}{\sqrt{N}} \left[ f(\xi_n^i) - f(\zeta_n^i) \right] \mid \xi_{n-1} \right)$$

For  $n = 0$ , the above conditional expectation is a simple expectation w.r.t. the initial random variables  $\xi_0$  and its independent copy  $\zeta_0$ . This yields the formula

$$V_n^N(f) = \mathbb{E} \left( M_{n,N}^{(N)}(f) \mid \xi_{n-1} \right)$$



with the terminal value  $M_{n,N}^{(N)}(f)$  of the martingale sequence defined below

$$i \in \{1, \dots, N\} \mapsto M_{n,i}^{(N)}(f) := \sum_{j=1}^i \frac{1}{\sqrt{N}} [f(\xi_n^j) - f(\zeta_n^j)]$$

By a direct application of Kintchine's inequality, we prove the following proposition

**Proposition 3.6.** *For any  $N \geq 1$ ,  $m \geq 1$ ,  $n \geq 0$  and any test function  $f$  we have the almost sure estimate*

$$\mathbb{E} \left( |V_n^N(f)|^m \mid \xi_{n-1} \right)^{\frac{1}{m}} \leq b(m) \operatorname{osc}(f) \quad (3.18)$$

The proof of the  $\mathbb{L}_m$ -mean errors stated in proposition 3.4 follows the same line of arguments as those developed in section 3.3.3. Once again, we use the decomposition

$$\begin{aligned} \eta_{n+1}^N(f) - \eta_{n+1}(f) &= [\eta_{n+1}^N(f) - \Phi_{n+1}(\eta_n^N)(f)] + [\Phi_{n+1}(\eta_n^N)(f) - \Phi_{n+1}(\eta_n)(f)] \\ &= \frac{1}{\sqrt{N}} V_{n+1}^N(f) + [\Phi_{n+1}(\eta_n^N)(f) - \Phi_{n+1}(\eta_n)(f)] \end{aligned}$$

and the fact that

$$|\Phi_{n+1}(\eta_n^N)(f) - \Phi_{n+1}(\eta_n)(f)| \leq \delta(G_n) \times |\eta_n^N(\tilde{f}_n) - \eta_n(\tilde{f}_n)|$$

with the function  $\tilde{f}_n$  defined in (3.13), to prove that

$$|\eta_{n+1}^N(f) - \eta_{n+1}(f)| \leq \frac{1}{\sqrt{N}} |V_{n+1}^N(f)| + \delta(G_n) \times |\eta_n^N(\tilde{f}_n) - \eta_n(\tilde{f}_n)|$$

This implies

$$\begin{aligned} \mathbb{E} \left( |\eta_{n+1}^N(f) - \eta_{n+1}(f)|^m \right)^{\frac{1}{m}} &\leq \frac{1}{\sqrt{N}} \mathbb{E} \left( |V_{n+1}^N(f)|^m \right)^{\frac{1}{m}} \\ &\quad + \delta(G_n) \times \mathbb{E} \left( |\eta_n^N(\tilde{f}_n) - \eta_n(\tilde{f}_n)|^m \right)^{\frac{1}{m}} \\ &\leq \frac{b(m)}{\sqrt{N}} + \delta(G_n) \times \mathbb{E} \left( |\eta_n^N(\tilde{f}_n) - \eta_n(\tilde{f}_n)|^m \right)^{\frac{1}{m}} \end{aligned}$$

The end of the proof can be conducted using a simple induction on the time parameter.

### 3.4.3. Proof of the bias estimates

The proof of the bias stated in proposition 3.4 needs a little bit more work. Using decomposition (3.14), we have

$$\Phi_{n+1}(\eta_n^N)(f) - \Phi_{n+1}(\eta_n)(f) = \frac{1}{1 - [1 - \eta_n^N(\tilde{g}_n)]} [\eta_n^N(\tilde{f}_n) - \eta_n(\tilde{f}_n)]$$

Using the fact that

$$\frac{1}{1-u} = 1 + u + \frac{u^2}{1-u},$$

we obtain

$$\frac{1}{1 - [1 - \eta_n^N(\tilde{g}_n)]} = 1 + [1 - \eta_n^N(\tilde{g}_n)] + \frac{[1 - \eta_n^N(\tilde{g}_n)]^2}{\eta_n^N(\tilde{g}_n)}.$$

Now we have

$$\begin{aligned}
& \mathbb{E}([\eta_{n+1}^N(f) - \eta_{n+1}(f)]) \\
&= \mathbb{E}(\Phi_{n+1}(\eta_n^N)(f)) - \Phi_{n+1}(\eta_n)(f) \\
&= \mathbb{E}([\eta_n^N(\tilde{f}_n) - \eta_n(\tilde{f}_n)]) + \mathbb{E}([\eta_n^N(\tilde{f}_n) - \eta_n(\tilde{f}_n)] \times [1 - \eta_n^N(\tilde{g}_n)]) \\
&\quad + \mathbb{E}\left([\eta_n^N(\tilde{f}_n) - \eta_n(\tilde{f}_n)] \times \frac{[1 - \eta_n^N(\tilde{g}_n)]^2}{\eta_n^N(\tilde{g}_n)}\right)
\end{aligned}$$

Using proposition 3.4, we have

$$\begin{aligned}
& |\mathbb{E}([\eta_n^N(\tilde{f}_n) - \eta_n(\tilde{f}_n)] \times [1 - \eta_n^N(\tilde{g}_n)])| \\
&\leq \mathbb{E}\left([\eta_n^N(\tilde{f}_n) - \eta_n(\tilde{f}_n)]^2\right)^{1/2} \times \mathbb{E}\left([1 - \eta_n^N(\tilde{g}_n)]^2\right)^{1/2} \leq c(n)/N
\end{aligned}$$

and in much the same way

$$\left| \mathbb{E}\left([\eta_n^N(\tilde{f}_n) - \eta_n(\tilde{f}_n)] \times \frac{[1 - \eta_n^N(\tilde{g}_n)]^2}{\eta_n^N(\tilde{g}_n)}\right) \right| \leq c(n)/(N\sqrt{N})$$

for some finite constant  $c(n) < \infty$ . This implies

$$|\mathbb{E}([\eta_{n+1}^N(f) - \eta_{n+1}(f)])| \leq |\mathbb{E}([\eta_n^N(\tilde{f}_n) - \eta_n(\tilde{f}_n)])| + c(n)/N$$

The end of the proof of the bias estimates stated in proposition 3.4 is now clear.

### 3.5. Central limit theorems

Let us come back to the interpretation of the mean field particle model as a stochastic perturbation of the limiting system. This perturbation model is summarized by the pair of formulae

$$\begin{aligned}
\eta_{n+1}^N &= \Phi_{n+1}(\eta_n^N) + \frac{1}{\sqrt{N}} V_{n+1}^N \\
\eta_{n+1} &= \Phi_{n+1}(\eta_n)
\end{aligned}$$

with the centered random measures  $V_n^N$  defined below

$$V_n^N(x) = \frac{1}{\sqrt{N}} \sum_{1 \leq i \leq N} [1_{\xi_n^i}(x) - K_{n, \eta_{n-1}^N}(\xi_{n-1}^i, x)]$$

These quantities can be thought as a stochastic perturbation of the limiting system, while the quantity  $\frac{1}{\sqrt{N}}$  is the usual precision of Monte Carlo approximation models.

Our next objective is to analyze the local sampling mean error variance associated with the particle model. To clarify the presentation, we slightly abuse the notation, and we set

$$x \in E \mapsto K_{n+1, \eta_n^N} \left( [f - K_{n+1, \eta_n^N}(f)]^2 \right) (x)$$

the function on  $E$  defined for any  $x$  by

$$K_{n+1, \eta_n^N} \left( [f - K_{n+1, \eta_n^N}(f)]^2 \right) (x) = \sum_x K_{n+1, \eta_n^N}(x, y) [f(y) - K_{n+1, \eta_n^N}(f)(x)]^2$$

**Lemma 3.7.** *For any  $N \geq 1$ , any  $n \geq 0$  and any functions  $f, g$  on  $E$  we have*

$$\begin{aligned} \mathbb{E}(V_n^N(f) | \xi_{n-1}) &= 0 \\ \mathbb{E}(V_n^N(f)^2 | \xi_{n-1}) &= \eta_{n-1}^N K_{n, \eta_{n-1}^N} \left( [f - K_{n, \eta_{n-1}^N}(f)]^2 \right) \end{aligned}$$

**Proof:**

The first assertion is immediate. Using the fact that

$$\begin{aligned} &\mathbb{E} \left( [f(\xi_{n+1}^i) - K_{n+1, \eta_n^N}(f)(\xi_n^i)]^2 | \xi_n \right) \\ &= \sum_x K_{n+1, \eta_n^N}(\xi_n^i, x) [f(x) - K_{n+1, \eta_n^N}(f)(\xi_n^i)]^2 := K_{n+1, \eta_n^N} \left( [f - K_{n+1, \eta_n^N}(f)]^2 \right) (\xi_n^i) \end{aligned}$$

we find that

$$\begin{aligned} &\mathbb{E}(V_n^N(f)^2 | \xi_{n-1}) \\ &= \frac{1}{N} \sum_{1 \leq i \leq N} \mathbb{E} \left( [f(\xi_{n+1}^i) - K_{n+1, \eta_n^N}(f)(\xi_n^i)]^2 | \xi_n \right) \\ &= \frac{1}{N} \sum_{1 \leq i \leq N} K_{n+1, \eta_n^N} \left( [f - K_{n+1, \eta_n^N}(f)]^2 \right) (\xi_n^i) := \eta_n^N K_{n+1, \eta_n^N} \left( [f - K_{n+1, \eta_n^N}(f)]^2 \right) \end{aligned}$$

This ends the proof of the lemma. ■

The next theorem indicates that these random perturbations behave asymptotically as Gaussian random perturbations. The details of the proof of this functional central limit theorem can be found in [9].

**Theorem 3.8.** *For any fixed time horizon  $n \geq 0$ , the sequence of random fields  $V_n^N$  converges in law, as the number of particles  $N$  tends to infinity, to a sequence of independent, Gaussian and centered random fields  $V_n$ ; with, for any  $f$ , and  $n \geq 0$ ,*

$$\mathbb{E}(V_n(f)^2) = \eta_{n-1} K_{n, \eta_{n-1}} ([f - K_{n, \eta_{n-1}}(f)]^2). \quad (3.19)$$

This fluctuation theorem can be used to analyze the fluctuations of the occupation measures  $\eta_n^N$  around their limiting value  $\eta_n$  (cf. for instance [9, 10], and the references therein). The extension of these results to general mean field models can be found in [13].

**Theorem 3.9.** *For any  $n \geq 0$  and any function  $f_n$  on  $E$  s.t.  $\eta_n(f_n) = 0$ , we have the following convergence in law*

$$\lim_{N \rightarrow \infty} \sqrt{N} \eta_n^N(f_n) = \sum_{q=0}^n V_q(\bar{Q}_{q,n}(f_n))$$

**Proof:**

We consider the following matrices:

$$\forall 0 \leq p \leq n \quad \bar{Q}_{p,n}(x, y) = \frac{Q_{p,n}(x, y)}{\eta_p Q_{p,n}(1)} \quad \text{with} \quad \eta_p Q_{p,n}(1) = \sum_{x, y} \eta_p(x) Q_{p,n}(x, y)$$

By construction, we have that

$$\eta_p \bar{Q}_{p,n} = \eta_n \quad (3.20)$$

We also observe that

$$\eta_p Q_{p,n}(1) = \frac{\gamma_p Q_{p,n}(1)}{\gamma_p(1)} = \frac{\gamma_n(1)}{\gamma_p(1)} = \prod_{p \leq q < n} \eta_q(G_q)$$

from which we prove that  $\bar{Q}_{p,n}$  is a well defined semigroup of matrices, in the sense that

$$\begin{aligned} \bar{Q}_{p,n}(x, y) &= Q_{p,n}(x, y) \prod_{p \leq q < n} \eta_q(G_q) \\ &= \left[ \prod_{p \leq q < r} \eta_q(G_q) \right] \left[ \prod_{r \leq q < n} \eta_q(G_q) \right] (Q_{p,r} Q_{r,n})(x, y) = \bar{Q}_{p,r} \bar{Q}_{r,n}(x, y) \end{aligned}$$

The following easily checked decomposition is pivotal

$$\begin{aligned} \eta_p^N \bar{Q}_{p,n}(f) &= \eta_0^N \bar{Q}_{0,n}(f) + \sum_{q=1}^p [\eta_q^N \bar{Q}_{q,n}(f) - \eta_{q-1}^N \bar{Q}_{q-1,n}(f)] \\ &= \eta_0^N \bar{Q}_{0,n}(f) + \sum_{q=1}^p [\eta_q^N \bar{Q}_{q,n}(f) - \Phi_q(\eta_{q-1}^N) \bar{Q}_{q,n}(f)] \\ &\quad + \sum_{q=1}^p [\Phi_q(\eta_{q-1}^N) \bar{Q}_{q,n}(f) - \eta_{q-1}^N \bar{Q}_{q-1,n}(f)] \end{aligned}$$

Choosing  $f = f_n$  s.t.  $\eta_n(f_n) = 0$ , we find that

$$\sqrt{N} \eta_p^N \bar{Q}_{p,n}(f_n) = \sum_{q=0}^p V_q^N (\bar{Q}_{q,n}(f_n)) + R_{p,n}^N(f_n)$$

with the remainder term

$$R_{p,n}^N(f_n) := \sqrt{N} \sum_{q=1}^p [\Phi_q(\eta_{q-1}^N) \bar{Q}_{q,n}(f_n) - \eta_{q-1}^N \bar{Q}_{q-1,n}(f_n)]$$

Using the fact that

$$\eta_{q-1}^N \bar{Q}_{q-1,n}(f_n) = \eta_{q-1}^N \bar{Q}_{q-1,q} \bar{Q}_{q,n}(f_n)$$

and

$$\Phi_q(\eta_{q-1}^N) \bar{Q}_{q,n}(f_n) = \frac{\eta_{q-1}^N Q_{q-1,q} \bar{Q}_{q,n}(f_n)}{\eta_{q-1}^N Q_{q-1,q}(1)} = \frac{\eta_{q-1}^N \bar{Q}_{q-1,q} \bar{Q}_{q,n}(f_n)}{\eta_{q-1}^N \bar{Q}_{q-1,q}(1)}$$

we prove that

$$[\Phi_q(\eta_{q-1}^N) \bar{Q}_{q,n}(f_n) - \eta_{q-1}^N \bar{Q}_{q-1,n}(f_n)] = \Phi_q(\eta_{q-1}^N) (\bar{Q}_{q,n}(f_n)) [1 - \eta_{q-1}^N \bar{Q}_{q-1,q}(1)]$$

By (3.20), we conclude that

$$\begin{aligned} &\Phi_q(\eta_{q-1}^N) (\bar{Q}_{q,n}(f_n)) [1 - \eta_{q-1}^N \bar{Q}_{q-1,q}(1)] \\ &= [\Phi_q(\eta_{q-1}^N) (\bar{Q}_{q,n}(f_n)) - \Phi_q(\eta_{q-1}^N) (\bar{Q}_{q,n}(f)))] [\eta_{q-1} \bar{Q}_{q-1,q}(1) - \eta_{q-1}^N \bar{Q}_{q-1,q}(1)] \end{aligned}$$

and therefore  $R_{p,n}^N(f_n) \rightarrow_{N \rightarrow \infty} 0$ , almost surely. The end of the proof is now a direct consequence of theorem 3.8. ■

## REFERENCES

- [1] D. Aldous and U. Vazirani. Go with the winners algorithms. In *Proc. 35th Symp. Foundations of Computer Sci.*, pp. 492-501, 1994.
- [2] R. Assaraf, M. Caffarel et A. Khelif. Diffusion Monte Carlo methods with a fixed number of walkers, *Phys. Rev. E*, vol. 61, pp. 4566-4575, 200.
- [3] J. Carpenter, P. Clifford and P. Fearnhead. An improved particle filter for non-linear problems. *IEE Proceedings F*, vol. 146, pp. 2-7, 1999.
- [4] R. Cerf. Asymptotic convergence of genetic algorithms. *Adv. Appl. Probab.*, vol. 30, pp. 521-550, 1998.
- [5] F. Cérou, P. Del Moral and A. Guyader. A non asymptotic variance theorem for unnormalized Feynman-Kac particle models. *Ann. Inst. H. Poincaré. Probab. Statist. Volume 47, Number 3*, 629-649, 2011.
- [6] F. Cérou, P. Del Moral, F. Le Gland and P. Lézaud. Genealogical models in entrance times rare event analysis. *Alea*, Latin American Journal of Probability And Mathematical Statistics 2006.
- [7] N. Chopin. A sequential particle filter method for static models. *Biometrika*, vol. 89, pp. 539-552, 2002.
- [8] J. E. David, W. D. Michael. Parallel tempering: Theory, applications, and new perspectives, *Phys. Chem. Chem. Phys.*, 7, 3910, 2005.
- [9] P. Del Moral. *Feynman-Kac Formulae: Genealogical and Interacting Particle Systems with Applications*, New York: Springer-Verlag, 2004.
- [10] P. Del Moral. *Mean Field Simulation for Monte Carlo Integration*, Chapman-Hall/CRC Press Monographs on Statistics & Applied Probability (2013).
- [11] P. Del Moral, A. Doucet and A. Jasra. Sequential Monte Carlo samplers. *J. Royal Statist. Soc. B*, vol. 68, pp. 411-436, 2006.
- [12] P. Del Moral, A. Doucet and A. Jasra. On Adaptive Resampling Procedures for Sequential Monte Carlo Methods *Bernoulli*, Vol. 18, No. 1, pp. 252-278 (2012).
- [13] P. Del Moral, and E. Rio. Concentration inequalities for mean field particle models. *Ann. Appl. Probab. Volume 21, Number 3*, 1017-1052 2011.
- [14] P. Del Moral and L. Miclo. Annealed Feynman-Kac models. *Comm. Math. Phys.*, vol. 235, pp. 191-214, 2003.
- [15] P. Del Moral and Alice Guionnet. On the stability of measure valued processes with applications to filtering. *C. R. Acad. Sci. Paris Sér. I Math.*, vol. 329, pp. 429-434, 1999.
- [16] P. Del Moral and A. Doucet. Particle motions in absorbing medium with hard and soft obstacles. *Stochastic Anal. Appl.*, vol. 22, pp. 1175-1207, 2004.
- [17] P. Del Moral and L. Miclo. Particle approximations of Lyapunov exponents connected to Schrödinger operators and Feynman-Kac semigroups. *ESAIM: Probability and Statistics*, no. 7, pp. 171-208, 2003.
- [18] P. Del Moral and L. Miclo. Genealogies and Increasing Propagations of Chaos for Feynman-Kac and Genetic Models. *Annals of Applied Probability*, vol. 11, No. 4, pp. 1166-1198 2001.
- [19] A. Doucet, J.F.G. de Freitas and N.J. Gordon (eds.) *Sequential Monte Carlo Methods in Practice*. New York: Springer-Verlag, 2001.
- [20] A. Doucet and A.M. Johansen. A tutorial on particle filtering and smoothing: fifteen years later. in *Handbook of Nonlinear Filtering*, Eds. D. Crisan and B. Rozovsky, Cambridge University Press, 2009.
- [21] A. Doucet, S.J. Godsill and C. Andrieu. On sequential Monte Carlo sampling methods for Bayesian filtering. *Statistics and Computing*, vol. 10, pp. 197-208, 2000.
- [22] P. Fearnhead. Computational methods for complex stochastic systems: A review of some alternatives to MCMC. *Statistics and Computing*, vol. 18, pp. 151-171, 2008.
- [23] H.R. Kunsch. State-space and hidden Markov models. in *Complex Stochastic Systems* (eds. O. E. Barndorff-Nielsen, D. R. Cox and C. Kluppelberg), CRC Press, pp. 109-173, 2001.
- [24] C. J. Geyer. in *Computing Science and Statistics, Proceedings of the 23rd Symposium on the Interface*, American Statistical Association, New York, p. 156, 1991.
- [25] F. Giraud, P. Del Moral Non-Asymptotic Analysis of Adaptive and Annealed Feynman-Kac Particle Models, arXiv preprint arXiv:1209.5654 2013.
- [26] P. Glasserman, P. Heidelberger, P. Shahabuddin, and T. Zajic. Multilevel splitting for estimating rare event probabilities. *Operations Research*, vol. 47, pp. 585-600, 1999.
- [27] D.E. Goldberg. *Genetic Algorithms in Search, Optimization and Machine Learning*. Addison-Wesley, Reading, MA, 1989..
- [28] N.J. Gordon, D. Salmond and A.F.M. Smith. A novel approach to state estimation to nonlinear non-Gaussian state estimation. *IEE Proceedings F*, vol. 40, pp. 107-113, 1993.

- [29] P. Grassberger. Pruned-enriched Rosenbluth method: Simulations of  $\theta$  polymers of chain length up to 1 000 000. *Phys. Rev. E*, pp. 3682-3693, 1997.
- [30] B. Hajek. Cooling schedules for optimal annealing, *Math. Oper. Res.* 13 311-329, 1988.
- [31] J.H. Hetherington. Observations on the statistical iteration of matrices, *Phys. Rev. A*, vol. 30, pp. 2713-2719, 1984.
- [32] J.H. Holland. *Adaptation in Natural and Artificial Systems*. University of Michigan Press, Ann Arbor 1975.
- [33] A. M. Johansen, P. Del Moral, and A. Doucet. Sequential Monte Carlo Samplers for Rare Events. in *Proceedings of 6th International Workshop on Rare Event Simulation*, Bamberg, Germany, 2006.
- [34] G. Kitagawa. Monte Carlo filter and smoother for non-Gaussian nonlinear state space models. *J. Comp. Graph. Statist.*, vol. 5, pp. 1-25, 1996.
- [35] J.S. Liu. *Monte Carlo Strategies in Scientific Computing*. New York: Springer Verlag, 2001.
- [36] J.S. Liu and R. Chen. Sequential Monte-Carlo methods for dynamic systems. *J. Am. Stat. Assoc.*, vol. 93, pp. 1032-1044, 1998.
- [37] V. Melik-Alaverdian and M.P. Nightingale, Quantum Monte Carlo methods in statistical mechanics, *Internat. J. of Modern Phys. C*, vol. 10, pp. 1409-1418, 1999.
- [38] R.H. Swendsen, J.S. Wang. Replica Monte Carlo simulation of spin glasses. *Physical Review Letters* 57 : 2607-2609, 1986.