Foundations and Trends® in
Machine Learning
Vol. 3, Nos. 3–4 (2010) 225–389

DOI: 10.1561/2200000026

now
the essence of knowledge

On the Concentration Properties of Interacting Particle Processes

By Pierre Del Moral, Peng Hu and Liming Wu

Contents

Foundations and Trends® in
Machine Learning
Vol. 3, Nos. 3–4 (2010) 225–389

DOI: 10.1561/2200000026

now
the essence of knowledge

On the Concentration Properties of Interacting Particle Processes

Pierre Del Moral[1], Peng Hu[2] and Liming Wu[3]

[1] *INRIA, Bordeaux-Sud Ouest Center, INRIA, Bordeaux-Sud Ouest Center & Bordeaux Mathematical Institute, Université Bordeaux 1, 351, Cours de la Libération, Talence, 33405, France, pierre.del-moral@inria.fr*

[2] *INRIA, Bordeaux-Sud Ouest Center, INRIA, Bordeaux-Sud Ouest Center & Bordeaux Mathematical Institute, Université Bordeaux 1, 351, Cours de la Libération, Talence, 33405, France, peng.hu@inria.fr*

[3] *Academy of Mathematics and Systems Science & Laboratoire de Mathématiques, Université Blaise Pascal, 63177 AUBIERE, France, Institute of Applied Mathematics, AMSS, CAS, Siyuan Building, Beijing, China, & Université Blaise Pascal, Aubière 63177, France, wuliming@amt.ac.cn, Li-Ming.Wu@math.univ-bpclermont.fr*

Abstract

This monograph presents some new concentration inequalities for Feynman-Kac particle processes. We analyze different types of stochastic particle models, including particle profile occupation measures, genealogical tree based evolution models, particle free energies, as well as backward Markov chain particle models. We illustrate these results with a series of topics related to computational physics and biology, stochastic optimization, signal processing and Bayesian statistics, and many other probabilistic machine learning algorithms. Special emphasis is given to the stochastic modeling, and to the quantitative performance analysis of a series of advanced Monte Carlo methods, including particle filters, genetic type island models, Markov bridge models, and interacting particle Markov chain Monte Carlo methodologies.

1

Stochastic Particle Methods

1.1 Introduction

Stochastic particle methods have come to play a significant role in applied probability, numerical physics, Bayesian statistics, probabilistic machine learning, and engineering sciences.

They are increasingly used to solve a variety of problems, including nonlinear filtering equations, data assimilation problems, rare event sampling, hidden Markov chain parameter estimation, stochastic control problems and financial mathematics. To name a few, They are also used in computational physics for free energy computations, and Schrödinger operator's ground states estimation problems, as well as in computational chemistry for sampling the conformation of polymers in a given solvent.

To illustrate these methods, we start with a classical filtering example. We consider a Markov chain X_k taking values in $\mathbb{R}^d$, with prior transitions given by

$$\mathbb{P}(X_k \in dx_k \mid X_{k-1} = x_{k-1}) = p_k(x_k|x_{k-1})\, dx_k, \tag{1.1}$$

Using some slight abuse of Bayesian notation, the observations Y_k are $\mathbb{R}^{d'}$-valued random variables defined in terms of the likelihood functions

$$\mathbb{P}(Y_k \in dy_k \mid X_k = x_k) = p_k(y_k|x_k)\ dy_k, \tag{1.2}$$

In the above display, dx_k and dy_k stand for the Lebesgue measures in $\mathbb{R}^d$ and $\mathbb{R}^{d'}$. To compute the conditional distribution of the signal path sequence $(X_0, \ldots, X_n)$, given the observations $(Y_0, \ldots, Y_n)$, we can use the genealogical tree model associated with a genetic type interacting particle model. This genetic algorithm is defined with mutation transitions according to 1.1, and proportional selections with regard to (w.r.t.) the fitness functions 1.2. The occupation measures of the corresponding genealogical tree provides an approximation of the desired conditional distributions of the signal. More generally, for any function f on the path space we have

$$\lim_{N\uparrow\infty} \frac{1}{N} \sum_{1}^{N} f(line_n(i)) = \mathbb{E}(f(X_0, \ldots, X_n)|Y_0 = y_0, \ldots, Y_n = y_n) \tag{1.3}$$

where $\text{line}_n(i)$ stands for the $i-$th ancestral line of the genealogical tree, at time n.

More refined particle filters can be designed, including fixed parameter estimates in hidden Markov chain models, unbiased particle estimates of the density of the observation sequence, and backward smoothing models based on complete ancestral trees. Section 2 presents a more rigorous and detailed discussion on these topics.

Rigorous understanding of these new particle Monte Carlo methodologies leads to fascinating mathematics related to Feynman-Kac path integral theory and their interacting particle interpretations [17, 20, 38]. In the last two decades, this line of research has been developed by using methods from stochastic analysis of interacting particle systems and nonlinear semigroup models in distribution spaces, but it has also generated difficult questions that cannot be addressed without developing new mathematical tools.

Let us survey some of the important challenges that arise.

For numerical applications, it is essential to obtain nonasymptotic quantitative information on the convergence of the algorithms. For instance, in the filtering problem presented at beginning of this section,

it is important to quantify the performance of the empirical particle estimate in 1.3. Asymptotic theory, including central limit theorems, moderate deviations, and large deviations principles have clearly limited practical values. An overview of these asymptotic results in the context of mean field and Feynman-Kac particle models can be found in the series of articles [13, 28, 29, 33, 41, 43].

Furthermore, when solving a given concrete problem, it is important to obtain explicit nonasymptotic error bounds estimates to ensure that the stochastic algorithm is provably correct. While non asymptotic propagation of chaos results provides some insights on the bias properties of these models, it rarely provides useful effective convergence rates.

Last but not least, it is essential to analyze the robustness properties, and more particularly the uniform performance of particle algorithms w.r.t. the time horizon. By construction, these important questions are intimately related to the stability properties of complex nonlinear Markov chain semigroups associated with the limiting measure valued process. In the filtering example illustrated in this section, the limiting measure valued process is given by the so-called nonlinear filtering equation. In this context, the stability property of these equations ensures that the optimal filter will correct any erroneous initial conditions. This line of thought has been further developed in the articles [13, 31, 38, 40], and in the books [17, 20].

Without any doubt, one of the most powerful mathematical tools to analyze the deviations of Monte Carlo based approximations is the theory of empirical processes and measure concentration theory. In the last two decades, these new tools have become one of the most important steps forward in infinite dimensional stochastic analysis, advanced machine learning techniques, as well as in the development of a statistical non asymptotic theory.

In recent years, much effort has been devoted to describing the behavior of the supremum norm of empirical functionals around the mean value of the norm. For an overview of these subjects, we refer the reader to the seminal books of Pollard [81], Van der Vaart and Wellner [93], Ledoux and Talagrand [72], the remarkable articles by Giné [56], Ledoux [70, 71], and Talagrand [90, 91, 92], and the more

recent article by Adamczak [1]. The best constants in Talagrand's concentration inequalities were obtained by Klein and Rio [67]. In this article, the authors proved the functional version of Bennett's and Bernstein's inequalities for sums of independent random variables.

Two main difficulties we encountered in applying these concentration inequalities to interacting particle models are of different order:

First, all of the concentration inequalities developed in the literature on empirical processes still involve the mean value of the supremum norm empirical functionals. In practical situations, these tail style inequalities can only be used if we have some precise information on the magnitude of the mean value of the supremum norm of the functionals.

On the other hand, the range of application of the theory of empirical processes and measure concentration theory is restricted to independent random samples, or equivalently product measures, and more recently to mixing Markov chain models. In the reverse angle, stochastic particle techniques are not based on fully independent sequences, nor on Markov chain Monte Carlo principles, but on interacting particle samples combined with complex nonlinear Markov chain semigroups. More precisely, in addition to the fact that particle models are built sequentially using conditionally independent random samples, their respective conditional distributions are still random. Also, in a nonlinear way, they strongly depend on the occupation measure of the current population.

In summary, the concentration analysis of interacting particle processes requires the development of new stochastic perturbation style techniques to control the interaction propagation and the degree of independence between the samples.

Del Moral and Ledoux [36] extend empirical processes theory to particle models. In this work, the authors proved Glivenko-Cantelli and Donsker theorems under entropy conditions, as well as nonasymptotic exponential bounds for Vapnik-Cervonenkis classes of sets or functions. Nevertheless, in practical situations these non asymptotic results tend to be a little disappointing, with very poor constants that degenerate w.r.t. the time horizon.

The second most important result on the concentration properties of the mean field particle model is found in [40]. This article is only

concerned with the finite marginal model. The authors generalize the classical Hoeffding, Bernstein and Bennett inequalities for independent random sequences to interacting particle systems.

In this monograph, we survey some of these results, and we provide new concentration inequalities for interacting empirical processes. We emphasize that this review does not give a comprehensive treatment of the theory of interacting empirical processes. To name a few missing topics, we do not discuss large deviation principles w.r.t. the strong τ-topology, Donsker type fluctuation theorems, moderate deviation principles, and continuous time models. The first two topics are developed [17], the third one is developed in [32], the last one is still an open research subject.

Here, we emphasize a single stochastic perturbation method, with second-order expansion entering the stability properties of the limiting Feynman-Kac semigroups. The concentration results attained are probably not the best possible of their kind. We have chosen to strive for just enough generality to derive useful and *uniform concentration inequalities w.r.t. the time horizon*, without having to impose complex and often unnatural regularity conditions to squeeze them into the general theory of empirical processes.

Some of the results are borrowed from [40], and many others are new. This monograph should be complemented with the books and articles [17, 20, 31, 44]. A very basic knowledge in statistics and machine learning theory will be useful, but not necessary. Good backgrounds in Markov chain theory and in stochastic semigroup analysis are necessary.

We have done our best to give a self-contained presentation, with detailed proofs. However, we assume some familiarity with Feynman-Kac models, and basic facts on the theory of Markov chains on abstract state spaces. Only in subsection 4.6.1, have we skipped the proof of some tools from convex analysis. We hope that the essential ideas are still accessible to the readers.

It is clearly not the scope of this monograph to give an exhaustive list of references to articles in computational physics, engineering sciences, and machine learning, presenting heuristic-like particle algorithms to solve a specific estimation problem. With a few exceptions, we have only provided references to articles with rigorous and well founded

mathematical treatments on particle models. We apologize in advance for possible errors, or for references that have been omitted due to the lack of accurate information.

This monograph grew from series of lectures the first author gave in the Computer Science and Communications Research Unit, of the University of Luxembourg in February and March 2011. They were reworked, with the addition of new material on the concentration of empirical processes for a course given at the Sino-French Summer Institute in Stochastic Modeling and Applications (CNRS-NSFC Joint Institute of Mathematics), held at the Academy of Mathematics and System Science, Beijing, in June 2011. The Summer Institute was ably organized by Fuzhou Gong, Ying Jiao, Gilles Pagès, and Mingyu Xu, and the members of the scientific committee, including Nicole El Karoui, Zhiming Ma, Shige Peng, Liming Wu, Jia-An Yan, and Nizar Touzi. The first author is grateful to them for giving to him the opportunity to experiment on a receptive audience with material not entirely polished.

In reworking the lectures, we have tried to resist the urge to push the analysis to general classes of mean field particle models, in the spirit of the recent joint article with E. Rio [40]. Our principal objective has been to develop just enough analysis to handle four types of Feynman-Kac interacting particle processes, namely, genetic dynamic population models, genealogical tree based algorithms, particle free energies, as well as backward Markov chain particle models. These application models do not exhaust the possible uses of the theory developed in these lectures.

1.2 A Brief Review on Particle Algorithms

Stochastic particle methods belong to the class of Monte Carlo methods. They can be thought of as a universal particle methodology for sampling complex distributions in highly dimensional state spaces.

We can distinguish two different classes of models, namely, diffusion type interacting processes, and interacting jump particle models. Feynman-Kac particle methods belongs to the second class of models, with rejection-recycling jump type interaction mechanisms. In contrast

to conventional acceptance-rejection type techniques, Feynman-Kac particle methods are equipped with an adaptive and interacting recycling strategy.

The common central feature of all the Monte Carlo particle methodologies developed so far is to solve discrete generation, or continuous time integro-differential equations in distribution spaces. The first heuristic-like description of these probabilistic techniques in mathematical physics goes back to the Los Alamos report [49], and the article by Everett and Ulam in 1948 [48], and the short article by Metropolis and Ulam [79], published in 1949.

In some instances, the flow of measures is dictated by the problem at hand. In advanced signal processing, the conditional distributions of the signal, given partial and noisy observations, are given by the so-called nonlinear filtering equation in distribution space (see for instance [15, 16, 17, 20, 38], and references therein).

Free energies and Schrödinger operator's ground states are given by the quasi-invariant distribution of a Feynman-Kac conditional distribution flow of non absorbed particles in absorbing media. We refer the reader to the articles by Cancès, Jourdain and Lelièvre [5], El Makrini, Jourdain and Lelièvre [46], Rousset [85], the pair of articles of Del Moral with Miclo [38, 39], with Doucet [19], and the book [17], and the references therein.

In mathematical biology, branching processes and infinite population models are also expressed by nonlinear parabolic type integro-differential equations. Further details on this subject can be found in the articles by Dawson and his co-authors [11, 12, 14], the works of Dynkin [45], and Le Gall [69], and more particularly the seminal book of Ethier and Kurtz [47], and the pioneering article by Feller [50].

In other instances, we formulate a given estimation problem in terms of a sequence of distributions with increasing complexity on state space models with increasing dimension. These stochastic evolutions can be related to decreasing temperature schedules in Boltzmann-Gibbs measures, multilevel decompositions for rare event excursion models on critical level sets, decreasing subsets strategies for sampling tail style distributions, and many other sequential

importance sampling plans. For a more thorough discussion on these models we refer the reader to [21].

From a purely probabilistic point of view, any flow of probability measures can be interpreted as the evolution of the laws of the random states of a Markov process. In contrast to conventional Markov chain models, the Markov transitions of these chains may depend on the distribution of the current random state. The mathematical foundations of these discrete generation models began in 1996 in [15] within the context of nonlinear filtering problems. Further analysis was developed in [38]. For a more thorough discussion on the origin and the performance analysis of these discrete generation models, we also refer the reader to the book [17], and the joint articles Del Moral with Guionnet [28, 29, 30, 31], and with Kouritzin [35].

The continuous time version of these nonlinear type Markov chain models take their origins from the 1960s, with the development of fluid mechanisms and statistical physics. We refer the reader to the pioneering works of McKean [61, 63], as well as the more recent treatments by Bellomo and Pulvirenti [3, 4], the series of articles by Graham and Méléard on interacting jump models [58, 59, 82], the articles by Méléard on Boltzmann equations [75, 76, 77, 78], and the lecture notes of Sznitman [89], and references therein.

In contrast to conventional Markov chain Monte Carlo techniques, these McKean type nonlinear Markov chain models can be thought of as perfect importance sampling strategies, in the sense that the desired *target measures coincide* at any time step with the law of the random states of a Markov chain. Unfortunately, as we mentioned above, the transitions of these chains depend on the distributions of their random states. Thus, they cannot be sampled without an additional level of approximation. One natural solution is to use a mean field particle interpretation model. These stochastic techniques belong to the class of stochastic population models, with free evolutions mechanisms, coupled with branching and/or adaptive interacting jumps. At any time step, the occupation measure of the population of individuals approximates the solution of the nonlinear equation, when the size of the system tends to ∞.

In genetic algorithms and sequential Monte Carlo literature, the reference free evolution model is interpreted as a reference sequence of twisted Markov chain samplers. These chains are used to perform the mutation/proposal transitions. As in conventional Markov chain Monte Carlo methods, the interacting jumps are interpreted as an acceptance-rejection transition, equipped with sophisticated interacting and adaptive recycling mechanism. In Bayesian statistics and engineering sciences, the resulting adaptive particle sampling model is often coined as a sequential Monte Carlo algorithm, genetic procedure, or simply a Sampling Importance Resampling method, mainly because it is based on importance sampling plans and online approximations of a flow of probability measures.

Since the 1960s, the adaptive particle recycling strategy has also been associated, in biology and engineering science, with several heuristic-like paradigms, with a proliferation of botanical names, depending on the application area in which they are considered: bootstrapping, switching, replenishing, pruning, enrichment, cloning, reconfigurations, resampling, rejuvenation, acceptance/rejection, spawning.

Of course, the idea of duplicating online better-fitted individuals and moving them one step forward to explore state-space regions is the basis of various stochastic search algorithms, such as:

Particle and bootstrap filters, Rao-Blackwell particle filters, sequential Monte Carlo methods, sequentially interacting Markov chain Monte Carlo methods, genetic type search algorithms, Gibbs cloning search techniques, interacting simulated annealing algorithms, sampling-importance resampling methods, quantum Monte Carlo walkers, adaptive population Monte Carlo sampling models, and many others evolutionary type Monte Carlo methods.

For a more detailed discussion on these models, with precise references, we refer the reader to the three books [17, 20, 44].

1.3 Feynman-Kac Path Integrals

Feynman-Kac measures represent the distribution of the paths of a Markov process, weighted by a collection of potential functions. These functional models encapsulate traditional changes of probability

measures, commonly used in importance sampling, posterior distributions in Bayesian statistics, and the optimal filter in nonlinear filtering problems.

These stochastic models are defined in terms of only two ingredients:

A Markov chain X_n, with Markov transition M_n on some measurable state spaces $(E_n, \mathcal{E}_n)$ with initial distribution η_0, and a sequence of $(0,1]$-valued potential functions G_n on the set E_n.

The Feynman-Kac path measure associated with the pairs (M_n, G_n) is the probability measure $\mathbb{Q}_n$ on the product state space

$$\mathbf{E}_n := (E_0 \times \ldots \times E_n)$$

defined by the following formula

$$d\mathbb{Q}_n := \frac{1}{\mathcal{Z}_n} \left\{ \prod_{0 \leq p < n} G_p(X_p) \right\} d\mathbb{P}_n \tag{1.4}$$

where $\mathcal{Z}_n$ is a normalizing constant and $\mathbb{P}_n$ is the distribution of the random paths

$$\mathbf{X_n} = (X_0, \ldots, X_n) \in \mathbf{E}_n$$

of the Markov process X_p from the origin $p = 0$ with initial distribution η_0, up to the current time $p = n$. We also denote by

$$\Gamma_n = \mathcal{Z}_n \mathbb{Q}_n \tag{1.5}$$

its unnormalized version.

The prototype model we have in mind is the traditional particle absorbed Markov chain model

$$X_n^c \in E_n^c := E_n \cup \{c\} \xrightarrow{absorption\ \sim(1-G_n)} \widehat{X}_n^c \xrightarrow{exploration\ \sim M_{n+1}} X_{n+1}^c. \tag{1.6}$$

The chain X_n^c starts at some initial state X_0^c randomly chosen with distribution η_0. During the absorption stage, we set $\widehat{X}_n^c = X_n^c$ with probability $G_n(X_n)$, otherwise we put the particle in an auxiliary cemetery state $\widehat{X}_n^c = c$. When the particle $\widehat{X}_n^c$ is still alive (that is, if we have $\widehat{X}_n^c \in E_n$), it performs an elementary move $\widehat{X}_n^c \rightsquigarrow X_{n+1}^c$ according to the

Markov transition M_{n+1}. Otherwise, the particle is absorbed and we set $X_p^c = \widehat{X}_p^c = c$, for any time $p > n$.

If we let T be the first time $\widehat{X}_n^c = c$, then we have the Feynman-Kac representation formulae

$$\mathbb{Q}_n = \mathrm{Law}((X_0^c, \ldots, X_n^c) \mid T \geq n) \quad \text{and} \quad \mathcal{Z}_n = \mathrm{Proba}(T \geq n).$$

For a more thorough discussion on the variety of application domains of Feynman-Kac models, we refer the reader to Section 2.

We also denote by η_n and γ_n, the n-th time marginal of $\mathbb{Q}_n$ and Γ_n. It is a simple exercise to check that

$$\gamma_n = \gamma_{n-1} Q_n \quad \text{and} \quad \eta_{n+1} = \Phi_{n+1}(\eta_n) := \Psi_{G_n}(\eta_n) M_{n+1} \tag{1.7}$$

with the positive integral operator

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

and the Boltzmann-Gibbs transformation

$$\Psi_{G_n}(\eta_n)(dx) = \frac{1}{\eta_n(G_n)} G_n(x) \eta_n(dx). \tag{1.8}$$

In addition, the normalizing constants $\mathcal{Z}_n$ can be expressed in terms of the flow of marginal measures η_p, from the origin $p = 0$ up to the current time n, with the following multiplicative formulae:

$$\mathcal{Z}_n := \gamma_n(\mathbb{1}) = \mathbb{E}\left(\prod_{0 \leq p < n} G_p(X_p) \right) = \prod_{0 \leq p < n} \eta_p(G_p). \tag{1.9}$$

This multiplicative formula is easily checked using the induction

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

The abstract formulae discussed above are more general than they may appear. For instance, they can be used to analyze, without further work, path spaces models, including historical processes or transition space models, as well as finite excursion models. These functional models also encapsulate quenched Feynman-Kac models, Brownian type bridges and linear Gaussian Markov chains conditioned on starting and end points.

For a more thorough discussion on these path space models, we refer the reader to subsections 2.4 and 2.6, Chapters 11–12 in the book [17], as well as to the Section 2, in this monograph.

When the Markov transitions M_n are absolutely continuous with respect to some measures λ_n on E_n, and for any $(x, y) \in (E_{n-1} \times E_n)$ we have

$$H_n(x, y) := \frac{dM_n(x, .)}{d\lambda_n}(y) > 0. \tag{1.10}$$

We also have the following backward formula

$$\mathbb{Q}_n(d(x_0, \ldots, x_n)) = \eta_n(dx_n) \prod_{q=1}^{n} \mathbb{M}_{q,\eta_{q-1}}(x_q, dx_{q-1}) \tag{1.11}$$

with the collection of Markov transitions defined by

$$\mathbb{M}_{n+1,\eta_n}(x, dy) \propto G_n(y) H_{n+1}(y, x) \eta_n(dy). \tag{1.12}$$

The proof of this formula is postponed to subsection 3.2.

Before launching into the description of the particle approximation of these models, we end this subsection with some connexions between discrete generation Feynman-Kac models and more conventional continuous time models arising in physics and scientific computing.

The Feynman-Kac models presented above play a central role in the numerical analysis of certain partial differential equations, offering a natural way to solve these functional integral models by simulating random paths of stochastic processes. These Feynman-Kac models were originally presented by Mark Kac in 1949 [66] for continuous time processes.

These continuous time models are used in molecular chemistry and computational physics to calculate the ground state energy of some Hamiltonian operators associated with some potential function V describing the energy of a molecular configuration (see, for instance, [5, 17, 39, 46, 85], and references therein). To better connect these partial differential equation models with (1.4), let us assume that $M_n(x_{n-1}, dx_n)$ is the Markov probability transition $X_n = x_n \rightsquigarrow X_{n+1} = x_{n+1}$ coming from a discretization in time $X_n = X'_{t_n}$ of a continuous time E-valued Markov process X'_t on a given time mesh $(t_n)_{n \geq 0}$ with

a given time step $(t_n - t_{n-1}) = \Delta t$. For potential functions of the form $G_n = e^{-V\Delta t}$, the measures $\mathbb{Q}_n \simeq_{\Delta t \to 0} \mathbb{Q}_{t_n}$ represent the time discretization of the following distribution:

$$d\mathbb{Q}_t = \frac{1}{\mathcal{Z}_t} \exp\left(-\int_0^t V(X'_s)ds\right) d\mathbb{P}_t^{X'}$$

where $\mathbb{P}_t^{X'}$ stands for the distribution of the random paths $(X'_s)_{0\leq s\leq t}$ with a given infinitesimal generator L. The marginal distributions γ_t at time t of the unnormalized measures $\mathcal{Z}_t\ d\mathbb{Q}_t$ are the solution of the so-called imaginary time Schrödinger equation, given in weak formulation on sufficiently regular function f by the following integro–differential equation

$$\frac{d}{dt}\ \gamma_t(f) := \gamma_t(L^V(f)) \quad \text{with} \quad L^V = L - V.$$

The errors introduced by the discretization of the time are well understood for regular models; we refer the interested reader to [34, 42, 68, 80] in the context of nonlinear filtering.

1.4 Interacting Particle Systems

Our aim here is to design an interacting particle approximation of the Feynman-Kac measures introduced in the previous subsection. These particle methods can be interpreted in different ways, depending on the application domain in which they are considered.

In the filtering example presented at the beginning of this monograph, these particle algorithms can be seen as a stochastic adaptive fixed approximation of the filtering equations. From a purely statistical point of view, these algorithms can also be seen as a sophisticated acceptance-rejection technique with an interacting recycling transition.

The particle model is defined as follows:

We start with a population of N candidate possible solutions $(\xi_0^1, \ldots, \xi_0^N)$ randomly chosen w.r.t. some distribution η_0.

The coordinates ξ_0^i are also called individuals or phenotypes, with $1 \leq N$. The random evolution of the particles is decomposed into two main steps : the free exploration and the adaptive selection transition.

During the updating-selection stage, multiple individuals in the current population $(\xi_n^1, \ldots, \xi_n^N)$ at time $n \in \mathbb{N}$ are stochastically selected based on the fitness function G_n. In practice, we choose a random proportion B_n^i of an existing solution ξ_n^i in the current population with a mean value $\propto G_n(\xi_n^i)$ to breed a brand new generation of "improved" solutions $(\widehat{\xi}_n^1, \ldots, \widehat{\xi}_n^N)$. For instance, for every index i, with a probability $\epsilon_n G_n(\xi_n^i)$, we set $\widehat{\xi}_n^i = \xi_n^i$, otherwise we replace ξ_n^i with a new individual $\widehat{\xi}_n^i = \xi_n^j$ randomly chosen from the whole population with a probability proportional to $G_n(\xi_n^j)$. The parameter $\epsilon_n \geq 0$ is a tuning parameter that must satisfy the constraint $\epsilon_n G_n(\xi_n^i) \leq 1$, for every $1 \leq i \leq N$. During the prediction-mutation stage, every selected individual $\widehat{\xi}_n^i$ moves to a new solution $\xi_{n+1}^i = x$ randomly chosen in E_{n+1}, with a distribution $M_{n+1}(\widehat{\xi}_n^i, dx)$.

If we interpret the updating-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 $\widehat{\xi}_{n-1}^i \longrightarrow \xi_n^i$ evolves to a new location ξ_n^i, we can interpret $\widehat{\xi}_{n-1}^i$ as the parent of ξ_n^i. Looking backwards in time and recalling that the particle $\widehat{\xi}_{n-1}^i$ has selected a site ξ_{n-1}^j in the configuration at time $(n-1)$, we can interpret this site ξ_{n-1}^j as the parent of $\widehat{\xi}_{n-1}^i$ and therefore as the ancestor denoted $\xi_{n-1,n}^i$ at level $(n-1)$ of ξ_n^i. Running backwards in time we may trace the whole ancestral line as

$$\xi_{0,n}^i \longleftarrow \xi_{1,n}^i \longleftarrow \cdots \longleftarrow \xi_{n-1,n}^i \longleftarrow \xi_{n,n}^i = \xi_n^i. \tag{1.13}$$

Most of the terminology we have used is drawn from filtering and genetic evolution theories.

In the filtering example presented in the subsection 1.1, the former particle model is dictated by the two steps prediction-updating learning equations of the conditional distributions of the signal process X_k, given some noisy and partial observations Y_k. In this setting, the potential functions represent the likelihood function of the current observation, while the free exploration transitions are related to the Markov transitions of the signal process. More formally, using the notation we used in example (1.2), we have:

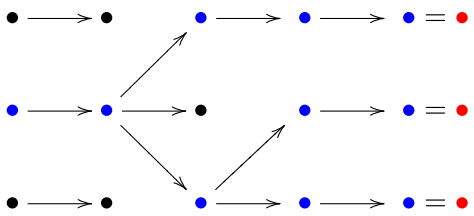
$$dp_k(x_k|x_{k-1}) = M_k(x_{k-1}|x_k)$$

and

$$p_k(y_k|x_k) = G_k(x_k).$$

In biology, the mutation-selection particle model presented above is used to mimic genetic evolutions of biological organisms and, more generally, natural evolution processes. For instance, in gene analysis, each population of individuals represents a chromosome and each individual particle is called a gene. In this setting the fitness potential function is usually time-homogeneous and it represents the quality and the adaptation potential value of the set of genes in a chromosome [62]. These particle algorithms are also used in population analysis to model changes in the structure of population in time and in space.

The different types of particle approximation measures associated with the genetic type particle model described above are summarized in the following synthetic picture corresponding to the case $N = 3$.



In the next four subsections we give an overview of the four particle approximation measures that can be extracted from the interacting population evolution model described above. We also provide some basic formulation of the concentration inequalities that will be treated in greater detail later. As a service to the reader we also provide precise pointers to their location within each section of the monograph.

We have already mentioned that the proofs of these results are quite subtle. In the further development of the next subsections, c_1 *stands for a finite constant related to* the bias *of the particle model, while* c_2 *is related to* the variance *of the scheme. The value of these constants may vary from one line to another, but in all the situations they do not depend on the time parameter.*

The precise form of the constants in these exponential inequalities depends on the contraction properties of Feynman-Kac flows. Our

stochastic analysis requires us to combine the stability properties of the nonlinear semigroup of the Feynman-Kac distribution flow η_n, with the deep convergence results of empirical processes theory associated with interacting random samples.

1.4.1 Last Population Models

The occupation measures of the current population, represented by the red dots in the above figure

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

converge to the n-th time marginals η_n of the Feynman-Kac measures $\mathbb{Q}_n$. We shall measure the performance of these particle estimates through several concentration inequalities, with a special emphasis on uniform inequalities w.r.t. the time parameter. Our results will basically be stated as follows.

1) For any time horizon $n \geq 0$, any bounded function f, any $N \geq 1$, and for any $x \geq 0$, the probability of the event

$$[\eta_n^N - \eta_n](f) \leq \frac{c_1}{N}(1 + x + \sqrt{x}) + \frac{c_2}{\sqrt{N}}\sqrt{x}$$

is greater than $1 - e^{-x}$.

We have already mentioned one important consequence of these uniform concentration inequalities for time homogeneous Feynman-Kac models. Under some regularity conditions, the flow of measures η_n tends to some fixed point distribution η_∞, in the sense that

$$\|\eta_n - \eta_\infty\|_{\mathrm{tv}} \leq c_3 \; e^{-\delta n} \tag{1.14}$$

for some finite positive constants c_3 and δ. In the above display $\|\nu - \mu\|_{tv}$ stands for the total variation distance. The connexions between these limiting measures and the top of the spectrum of Schrödinger operators is discussed in subsection 2.7.1. We also refer the reader to subsection 2.7.2 for a discussion on these quasi-invariant measures and Yaglom limits. Quantitative contraction theorems for Feynman-Kac semigroups are developed in subsection 3.4.2. As a direct consequence of the above inequalities, we find that for any $x \geq 0$, the

probability of the following events is greater than $1 - e^{-x}$:

$$[\eta_n^N - \eta_\infty](f) \leq \frac{c_1}{N}(1 + x + \sqrt{x}) + \frac{c_2}{\sqrt{N}}\sqrt{x} + c_3 e^{-\delta n}.$$

2) For any $x = (x_i)_{1 \leq i \leq d} \in E_n = \mathbb{R}^d$, we set $(-\infty, x] = \prod_{i=1}^d (-\infty, x_i]$ and we consider the repartition functions

$$F_n(x) = \eta_n(1_{(-\infty,x]}) \quad \text{and} \quad F_n^N(x) = \eta_n^N(1_{(-\infty,x]}).$$

The probability of the following event

$$\sqrt{N}\|F_n^N - F_n\| \leq c\sqrt{d(x+1)}$$

is greater than $1 - e^{-x}$, for any $x \geq 0$, for some universal constant $c < \infty$ that does not depend on the dimension, nor on the time parameter. In the above display $\|F\| = \sup_x |F(x)|$ stands for the uniform norm. Furthermore, under the stability properties (1.14), if we set

$$F_\infty(x) = \eta_\infty(1_{(-\infty,x]})$$

then, the probability of the following event

$$\|F_n^N - F_\infty\| \leq \frac{c}{\sqrt{N}}\sqrt{d(x+1)} + c_3 e^{-\delta n}$$

is greater than $1 - e^{-x}$, for any $x \geq 0$, for some universal constant $c < \infty$ that does not depend on the dimension.

For more precise statements, we refer the reader to Corollary 6.4, and Corollary 6.9, respectively.

The concentration properties of the particle measures η_n^N around their limiting values are developed in Section 6. In subsection 6.3, we design a stochastic perturbation analysis that allows us to enter the stability properties of the limiting Feynman-Kac semigroup. Finite marginal models are discussed in subsection 6.4.1. Subsection 6.4.2 is concerned with the concentration inequalities of interacting particle processes w.r.t. some collection of functions.

1.4.2 Particle Free Energy Models

Mimicking the multiplicative formula (1.9), we set

$$\mathcal{Z}_n^N = \prod_{0 \leq p < n} \eta_p^N(G_p) \quad \text{and} \quad \gamma_n^N(dx) = \mathcal{Z}_n^N \times \eta_n^N(dx). \tag{1.15}$$

We have already mentioned that these rather complex particle models provide an unbiased estimate of the unnormalized measures. That is, we have that

$$\mathbb{E}\left(\eta_n^N(f_n)\prod_{0\leq p<n}\eta_p^N(G_p)\right)=\mathbb{E}\left(f_n(X_n)\prod_{0\leq p<n}G_p(X_p)\right). \tag{1.16}$$

The concentration properties of the *unbiased* particle free energies $\mathcal{Z}_n^N$ around their limiting values $\mathcal{Z}_n$ are developed in subsection 6.5. Our results will basically be stated as follows.

For any $N\geq 1$, and any $\epsilon\in\{+1,-1\}$, the probability of each of the following events

$$\frac{\epsilon}{n}\log\frac{\mathcal{Z}_n^N}{\mathcal{Z}_n}\leq\frac{c_1}{N}(1+x+\sqrt{x})+\frac{c_2}{\sqrt{N}}\sqrt{x}$$

is greater than $1-e^{-x}$. A more precise statement is provided in Corollary 6.14.

1.4.3 Genealogical Tree Model

The occupation measure of the N-genealogical tree model represented by the lines linking the blue dots converges as $N\to\infty$ to the distribution $\mathbb{Q}_n$

$$\lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\delta_{(\xi_{0,n}^i,\xi_{1,n}^i,\ldots,\xi_{n,n}^i)}=\mathbb{Q}_n. \tag{1.17}$$

Our concentration inequalities will basically be stated as follows. A more precise statement is provided in Corollary 6.5.

For any $n\geq 0$, any bounded function $\mathbf{f_n}$ on the path space $\mathbf{E_n}$, such that (s.t.) $\|\mathbf{f_n}\|\leq 1$, and any $N\geq 1$, the probability of each of the following events

$$\left[\frac{1}{N}\sum_{i=1}^{N}\mathbf{f_n}(\xi_{0,n}^i,\xi_{1,n}^i,\ldots,\xi_{n,n}^i)-\mathbb{Q}_n(\mathbf{f_n})\right]$$
$$\leq c_1\frac{n+1}{N}(1+x+\sqrt{x})+c_2\sqrt{\frac{(n+1)}{N}}\sqrt{x}$$

is greater than $1-e^{-x}$.

The concentration properties of genealogical tree occupation measures can be derived more or less directly from those of the current population models. This rather surprising assertion comes from the fact that the n-th time marginal η_n of a Feynman-Kac measure associated with a reference historical Markov process has the same form as in the measure (1.4). This equivalence principle between $\mathbb{Q}_n$ and the marginal measures are developed in subsection 3.2, dedicated to historical Feynman-Kac models.

Using these properties, we prove concentration properties for interacting empirical processes associated with genealogical tree models. Our concentration inequalities will basically be stated as follows. A more precise statement is provided in subsection 6.4.2. We let $\mathcal{F}_n$ be the set of product functions of cell indicators in the path space $\mathbf{E_n} = (\mathbb{R}^{d_0} \times \ldots, \times \mathbb{R}^{d_n})$, for some $d_p \geq 1$, $p \geq 0$. We also denote by η_n^N the occupation measure of the genealogical tree model. In this notation, the probability of the following event

$$\sup_{\mathbf{f_n} \in \mathcal{F}_n} |\eta_n^N(\mathbf{f_n}) - \mathbb{Q}_n(\mathbf{f_n})| \leq c(n+1)\sqrt{\frac{\sum_{0 \leq p \leq n} d_p}{N}(x+1)}$$

is greater than $1 - e^{-x}$, for any $x \geq 0$, for some universal constant $c < \infty$ that does not depend on the dimension.

1.4.4 Complete Genealogical Tree Models

Mimicking the backward model (1.11) and the above formulae, we set

$$\Gamma_n^N = \mathcal{Z}_n^N \times \mathbb{Q}_n^N \tag{1.18}$$

with

$$\mathbb{Q}_n^N(d(x_0, \ldots, x_n)) = \eta_n^N(dx_n) \prod_{q=1}^{n} \mathbb{M}_{q, \eta_{q-1}^N}(x_q, dx_{q-1}).$$

Notice that the computation of sums w.r.t. these particle measures are reduced to summations over the particle locations ξ_n^i. It is therefore natural to identify a population of individuals $(\xi_n^1, \ldots, \xi_n^N)$ at time n to the ordered set of indexes $\{1, \ldots, N\}$. In this case, the occupation

measures and the functions are identified with the following line and column vectors

$$\eta_n^N := \left[\frac{1}{N}, \ldots, \frac{1}{N}\right] \quad \text{and} \quad \mathrm{f}_n := \begin{pmatrix} f_n(\xi_n^1) \\ \vdots \\ f_n(\xi_n^N) \end{pmatrix}$$

and the matrices $\mathbb{M}_{n,\eta_{n-1}^N}$ by the $(N \times N)$ matrices

$$\mathbb{M}_{n,\eta_{n-1}^N} := \begin{pmatrix} \mathbb{M}_{n,\eta_{n-1}^N}(\xi_n^1, \xi_{n-1}^1) & \cdots & \mathbb{M}_{n,\eta_{n-1}^N}(\xi_n^1, \xi_{n-1}^N) \\ \vdots & \vdots & \vdots \\ \mathbb{M}_{n,\eta_{n-1}^N}(\xi_n^N, \xi_{n-1}^1) & \cdots & \mathbb{M}_{n,\eta_{n-1}^N}(\xi_n^N, \xi_{n-1}^N) \end{pmatrix} \tag{1.19}$$

with the (i,j)-entries

$$\mathbb{M}_{n,\eta_{n-1}^N}(\xi_n^i, \xi_{n-1}^j) = \frac{G_{n-1}(\xi_{n-1}^j) H_n(\xi_{n-1}^j, \xi_n^i)}{\sum_{k=1}^N G_{n-1}(\xi_{n-1}^k) H_n(\xi_{n-1}^k, \xi_n^i)}.$$

For instance, the $\mathbb{Q}_n$-integration of normalized additive linear functionals of the form

$$\mathbf{f}_n(x_0, \ldots, x_n) = \frac{1}{n+1} \sum_{0 \le p \le n} f_p(x_p) \tag{1.20}$$

is given the particle matrix approximation model

$$\mathbb{Q}_n^N(\mathbf{f}_n) = \frac{1}{n+1} \sum_{0 \le p \le n} \eta_n^N \mathbb{M}_{n,\eta_{n-1}^N} \mathbb{M}_{n-1,\eta_{n-2}^N} \ldots \mathbb{M}_{p+1,\eta_p^N}(f_p).$$

These type of additive functionals arise in the calculation of the sensitivity measures discussed in subsection 2.4.1.

The concentration properties of the particle measures $\mathbb{Q}_n^N$ around the Feynman-Kac measures $\mathbb{Q}_n$ are developed in subsection 6.6. Special emphasis is given to the additive functional models (1.20). In subsection 6.6.3, we extend the stochastic perturbation methodology developed in subsection 6.3 for time marginal models to the particle backward Markov chain associated with the random stochastic matrices (1.19). This technique allows us to enter not only the stability properties of the limiting Feynman-Kac semigroup, but also those of the particle backward Markov chain model.

Our concentration inequalities will basically be stated as follows. A more precise statement is provided in Corollary 6.20 and in Corollary 6.24.

For any $n \geq 0$, any normalized additive functional of the form (1.20), with $\max_{0 \leq p \leq n} \|f_p\| \leq 1$, and any $N \geq 1$, the probability of each of the following events

$$[\mathbb{Q}_n^N - \mathbb{Q}_n](\overline{\mathbf{f}}_n) \leq c_1 \frac{1}{N} \ (1 + x + \sqrt{x}) + c_2 \sqrt{\frac{x}{N(n+1)}}$$

is greater than $1 - e^{-x}$.

For any $a = (a_i)_{1 \leq i \leq d} \in E_n = \mathbb{R}^d$, we denote by C_a the cell

$$C_a := (-\infty, a] = \prod_{i=1}^{d} (-\infty, a_i]$$

and $\mathbf{f}_{a,n}$ the additive functional

$$\mathbf{f}_{a,n}(x_0, \ldots, x_n) = \frac{1}{n+1} \sum_{0 \leq p \leq n} 1_{(-\infty,a]}(x_p).$$

The probability of the following event

$$\sup_{a \in \mathbb{R}^d} |\mathbb{Q}_n^N(\mathbf{f}_{a,n}) - \mathbb{Q}_n(\mathbf{f}_{a,n})| \leq c \sqrt{\frac{d}{N}(x+1)}$$

is greater than $1 - e^{-x}$, for any $x \geq 0$, for some constant $c < \infty$ that does not depend on the dimension, nor on the time horizon.

Remark 1.1. One way to turn all of these inequalities into of Bernstein style concentration inequalities is as follows. For any exponential inequality of the form

$$\forall x \geq 0 \quad \mathbb{P}(X \leq ax + \sqrt{2bx} + c) \leq 1 - e^{-x}$$

for some non negative constants (a, b, c), we also have

$$\forall y \geq 0 \quad \mathbb{P}(X \leq y + c) \leq 1 - \exp\left(-\frac{y^2}{2\,(b + ay)}\right).$$

A proof of this result is provided in Lemma 4.13.

1.5 Basic Notation

Here we provide some background from stochastic analysis and integral operator theory, which we require for our proofs. Most of the results with detailed proofs can be located in the book [17] on Feynman-Kac formulae and interacting particle methods. Our proofs also contain cross-references to this rather well known material, so the reader may wish to skip this subsection and proceed directly to Section 2, which is dedicated to some application domains of Feynman-Kac models.

1.5.1 Integral Operators

We denote respectively by $\mathcal{M}(E)$, $\mathcal{M}_0(E)$, $\mathcal{P}(E)$, and $\mathcal{B}(E)$, the set of all finite signed measures on some measurable space $(E,\mathcal{E})$, the convex subset of measures with null mass, the set of all probability measures, and the Banach space of all bounded and measurable functions f equipped with the uniform norm $\|f\|$. We also denote by $\mathrm{Osc}_1(E)$, and by $\mathcal{B}_1(E)$ the set of $\mathcal{E}$-measurable functions f with oscillations $\mathrm{osc}(f) \leq 1$, and respectively with $\|f\| \leq 1$. We let

$$\mu(f) = \int \mu(dx) f(x)$$

be the Lebesgue integral of a function $f \in \mathcal{B}(E)$, with respect to a measure $\mu \in \mathcal{M}(E)$.

We recall that the total variation distance on $\mathcal{M}(E)$ is defined for any $\mu \in \mathcal{M}(E)$ by

$$\|\mu\|_{\mathrm{tv}} = \frac{1}{2} \sup_{(A,B)\in\mathcal{E}^2} (\mu(A) - \mu(B)).$$

We recall that a bounded integral operator M turned from a measurable space $(E,\mathcal{E})$ into an auxiliary measurable space $(F,\mathcal{F})$ is an operator $f \mapsto M(f)$ from $\mathcal{B}(F)$ into $\mathcal{B}(E)$ such that the functions

$$M(f)(x) := \int_F M(x,dy) f(y)$$

are $\mathcal{E}$-measurable and bounded, for any $f \in \mathcal{B}(F)$. A Markov kernel is a positive and bounded integral operator M with $M(1) = 1$. Given a pair

of bounded integral operators (M_1, M_2), we let $(M_1 M_2)$ represent the composition operator defined by $(M_1 M_2)(f) = M_1(M_2(f))$. For time homogeneous state spaces, we denote by $M^m = M^{m-1}M = MM^{m-1}$ the m-th composition of a given bounded integral operator M, with $m \geq 1$. A bounded integral operator M turned from a measurable space $(E, \mathcal{E})$ into an auxiliary measurable space $(F, \mathcal{F})$ also generates a dual operator

$$\mu(dx) \mapsto (\mu M)(dx) = \int \mu(dy) M(y, dx)$$

from $\mathcal{M}(E)$ into $\mathcal{M}(F)$ defined by $(\mu M)(f) := \mu(M(f))$. We also used the notation

$$K([f - K(f)]^2)(x) := K([f - K(f)(x)]^2)(x)$$

for some bounded integral operator K and some bounded function f.

We prefer to avoid unnecessary abstraction and technical assumptions, so we frame the standing assumption that all the test functions are in the unit sphere, and the integral operators, and all the random variables are sufficiently regular so that we are justified in computing integral transport equations, regular versions of conditional expectations, and so forth.

1.5.2 Contraction Coefficients

When the bounded integral operator M has a constant mass, that is, when $M(1)(x) = M(1)(y)$ for any $(x, y) \in E^2$, the operator $\mu \mapsto \mu M$ maps $\mathcal{M}_0(E)$ into $\mathcal{M}_0(F)$. In this situation, we let $\beta(M)$ be the Dobrushin coefficient of a bounded integral operator M defined by the formula

$$\beta(M) := \sup \ \{\mathrm{osc}(M(f)); \ f \in \mathrm{Osc}(F)\}.$$

Notice that $\beta(M)$ is the operator norm of M on $\mathcal{M}_0(E)$, and we have the equivalent formulations

$$\begin{aligned} \beta(M) &= \sup \{\|M(x, .) - M(y, .)\|_{\mathrm{tv}} \ ; \ (x, y) \in E^2\} \\ &= \sup_{\mu \in \mathcal{M}_0(E)} \|\mu M\|_{\mathrm{tv}} / \|\mu\|_{\mathrm{tv}}. \end{aligned}$$

A detailed proof of these well known formulae can be found in [17].

Given a positive and bounded potential function G on E, we also denote by Ψ_G the Boltzmann-Gibbs mapping from $\mathcal{P}(E)$ into itself, defined for any $\mu \in \mathcal{P}(E)$ by

$$\Psi_G(\mu)(dx) = \frac{1}{\mu(G)} G(x)\mu(dx).$$

For $[0,1]$-valued potential functions, we also mention that $\Psi_G(\mu)$ can be expressed as a non linear Markov transport equation

$$\Psi_G(\mu) = \mu S_{\mu,G} \tag{1.21}$$

with the Markov transitions

$$S_{\mu,G}(x,dy) = G(x)\delta_x(dy) + (1 - G(x))\Psi_G(\mu)(dy).$$

We notice that

$$\Psi_G(\mu) - \Psi_G(\nu) = (\mu - \nu)S_\mu + \nu(S_\mu - S_\nu)$$

and

$$\nu(S_\mu - S_\nu) = (1 - \nu(G))[\Psi_G(\mu) - \Psi_G(\nu)]$$

from which we find the formula

$$\Psi_G(\mu) - \Psi_G(\nu) = \frac{1}{\nu(G)}\ (\mu - \nu)S_\mu.$$

In addition, using the fact that

$$\forall (x,A) \in (E,\mathcal{E}) \quad S_\mu(x,A) \geq (1 - \|G\|)\Psi_G(\mu)(A)$$

we prove that $\beta(S_\mu) \leq \|G\|$ and

$$\|\Psi_G(\mu) - \Psi_G(\nu)\|_{\mathrm{tv}} \leq \frac{\|G\|}{\mu(G) \vee \nu(G)}\ \|\mu - \nu\|_{\mathrm{tv}}.$$

If we set $\Phi(\mu) = \Psi_G(\mu)M$, for some Markov transition M, then we have the decomposition

$$\Phi(\mu) - \Phi(\nu) = \frac{1}{\nu(G)}\ (\mu - \nu)S_\mu M \tag{1.22}$$

for any couple of measures ν, μ on E. From the previous discussion, we also find the following Lipschitz estimates

$$\|\Phi(\mu) - \Phi(\nu)\|_{\mathrm{tv}} \leq \frac{\|G\|}{\mu(G) \vee \nu(G)} \beta(M) \|\mu - \nu\|_{\mathrm{tv}}. \tag{1.23}$$

We end this subsection with an interesting contraction property of a Markov transition

$$M_G(x, dy) = \frac{M(x, dy)G(y)}{M(G)(x)} = \Psi_G(\delta_x M)(dy) \tag{1.24}$$

associated with a $]0,1]$-valued potential function G, with

$$g = \sup_{x,y} G(x)/G(y) < \infty. \tag{1.25}$$

It is easily checked that

$$\begin{aligned} |M_G(f)(x) - M_G(f)(y)| &= |\Psi_G(\delta_x M)(f) - \Psi_G(\delta_y M)(f)| \\ &\leq g \|\delta_x M - \delta_y M\|_{\mathrm{tv}} \end{aligned}$$

from which we conclude that

$$\beta(M_G) \leq g \beta(M). \tag{1.26}$$

1.5.3 Orlicz Norms and Gaussian Moments

We let $\pi_\psi[Y]$ be the Orlicz norm of an $\mathbb{R}$-valued random variable Y associated with the convex function $\psi(u) = e^{u^2} - 1$, and defined by

$$\pi_\psi(Y) = \inf \{a \in (0, \infty) : \mathbb{E}(\psi(|Y|/a)) \leq 1\}$$

with the convention $\inf_\emptyset = \infty$. Notice that

$$\pi_\psi(Y) \leq c \Longleftrightarrow \mathbb{E}(\psi(Y/c)) \leq 1.$$

For instance, the Orlicz norm of a Gaussian and centered random variable U, s.t. $E(U^2) = 1$, is given by $\pi_\psi(U) = \sqrt{8/3}$. We also recall that

$$\begin{aligned} \mathbb{E}(U^{2m}) &= b(2m)^{2m} := (2m)_m \, 2^{-m} \\ \mathbb{E}(|U|^{2m+1}) &\leq b(2m+1)^{2m+1} := \frac{(2m+1)_{(m+1)}}{\sqrt{m + 1/2}} \, 2^{-(m+1/2)} \end{aligned} \tag{1.27}$$

with $(q+p)_p := (q+p)!/q!$. The second assertion comes from the fact that

$$\mathbb{E}(U^{2m+1})^2 \leq \mathbb{E}(U^{2m})\mathbb{E}(U^{2(m+1)})$$

and therefore

$$\begin{aligned} b(2m+1)^{2(2m+1)} &= \mathbb{E}(U^{2m})\mathbb{E}(U^{2(m+1)}) \\ &= 2^{-(2m+1)}\ (2m)_m\ (2(m+1))_{(m+1)}. \end{aligned}$$

This formula is a direct consequence of the following decompositions

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

and

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

We also mention that

$$b(m) \leq b(2m). \tag{1.28}$$

Indeed, for even numbers $m = 2p$ we have

$$b(m)^{2m} = b(2p)^{4p} = \mathbb{E}(U^{2p})^2 \leq \mathbb{E}(U^{4p}) = b(4p)^{4p} = b(2m)^{2m}$$

and for odd numbers $m = (2p+1)$, we have

$$\begin{aligned} b(m)^{2m} &= b(2p+1)^{2(2p+1)} = \mathbb{E}(U^{2p})\mathbb{E}(U^{2(p+1)}) \\ &\leq \mathbb{E}\left((U^{2p})^{\frac{(2p+1)}{p}}\right)^{\frac{p}{2p+1}} \mathbb{E}\left((U^{2(p+1)})^{\frac{(2p+1)}{p+1}}\right)^{\frac{p+1}{2p+1}} \\ &= \mathbb{E}(U^{2(2p+1)}) = b(2(2p+1))^{2(2p+1)} = b(2m)^{2m}. \end{aligned}$$

2

Some Application Domains

2.1 Introduction

Particle methods have been used in physics, biology, and engineering science since the beginning of the 1950's. We refer the reader to [17] for a detailed survey on their application domains, and a list of precise bibliographic references.

Surprisingly, most of these interacting particle methods can be encapsulated in a single mathematical model. This model is given by the Feynman-Kac measures presented in subsection 1.3. The corresponding methodology developed in subsection 1.4 is termed quantum Monte Carlo methods in computational physics, genetic algorithms in computer sciences, and particle filters and/or sequential Monte Carlo methods in information theory, as well as in Bayesian statistics.

The mathematical foundations of these advanced interacting Monte Carlo methodologies are now fifteen years old [15]. Since this period, many descriptions and variants of these models have been published in applied probability, signal processing and Bayesian statistic literature.

For a detailed discussion on their application domains with a precise bibliography of who first did what when, we refer the reader to any of the following references [17, 20, 38, 44].

Here, we merely content ourselves in illustrating the rather abstract models (1.4) with the Feynman-Kac representations of 20 more or less well known conditional distributions, including three more recent applications related to island particle models, functional kinetic parameter derivatives, and gradient analysis of Markov semigroups.

The forthcoming series of examples, combined with their mean field particle interpretation models described in subsection 1.4, also illustrate the ability of the Feynman-Kac particle methodology to solve complex conditional distribution flows as well as their normalizing constants.

Of course, this selected list of applications does not attempt to be exhaustive. The topics selection is largely influenced by the personal taste of the authors. A complete description on how particle methods are applied in each application model area would of course require separate volumes, with precise computer simulations and comparisons with different types of particle models and other existing algorithms.

We also limit ourselves to describing the key ideas in a simple way, often sacrificing generality. Some applications are nowadays routine, and in this case we provide precise pointers to existing, more application-related, articles in the literature. Readers who wish to know more about some specific application of these particle algorithms are invited to consult the referenced papers.

One natural path of "easy reading" will probably be to choose a familiar or attractive application area and to explore some selected parts of the monograph in terms of this choice. Nevertheless, this advice must not be taken too literally. To see the impact of particle methods, it is essential to understand the full force of Feynman-Kac modeling techniques on various research domains. Upon doing so, the reader will have a powerful weapon for the discovery of new particle interpretation models. The principal challenge is to understand the theory well enough to reduce it to practice.

2.2 Boltzmann-Gibbs Measures

2.2.1 Interacting Markov Chain Monte Carlo (MCMC) Methods

Suppose we are given a sequence of target probability measures on some measurable state space E of the following form

$$\mu_n(dx) = \frac{1}{\mathcal{Z}_n} \left\{ \prod_{0 \leq p < n} h_p(x) \right\} \lambda(dx) \tag{2.1}$$

with some sequence of bounded nonnegative potential functions

$$h_n : x \in E \mapsto h_n(x) \in (0, \infty)$$

and some reference probability measure λ on E. In the above displayed formula, $\mathcal{Z}_n$ stands for a normalizing constant. We use the convention $\prod_{\emptyset} = 1$ and $\mu_0 = \lambda$.

We further assume that we have a dedicated Markov chain Monte Carlo transition M_n with prescribed target invariant measures μ_n, at any time step, in the sense that $\mu_n = \mu_n M_n$. Using the fact that

$$\mu_{n+1} = \Psi_{h_n}(\mu_n)$$

with the Boltzmann-Gibbs transformations defined in (1.8), we prove that

$$\mu_{n+1} = \mu_{n+1} M_{n+1} = \Psi_{h_n}(\mu_n) M_{n+1}$$

from which we conclude that

$$\mu_n(f) = \mathbb{E}\left(f(X_n) \prod_{0 \leq p < n} h_p(X_p) \right) \Big/ \mathbb{E}\left(\prod_{0 \leq p < n} h_p(X_p) \right)$$

with the reference Markov chain

$$\mathbb{P}(X_n \in dx \mid X_{n-1}) = M_n(X_{n-1}, dx).$$

In addition, we have

$$\mathcal{Z}_{n+1} = \int \left[\prod_{0 \leq p < n} h_p(x) \right] h_n(x) \mu(dx) = \mathcal{Z}_n \mu_n(h_n) = \prod_{0 \leq p \leq n} \mu_p(h_p).$$

We illustrate these rather abstract models with two applications related to probability restriction models and stochastic optimization simulated annealing type models, respectively. For a more thorough discussion on these interacting MCMC models, and related sequential Monte Carlo methods, we refer the reader to [17, 21, 22].

2.2.2 Probability Restrictions

If we choose Markov chain Monte Carlo type local moves

$$\mu_n = \mu_n M_n$$

with some prescribed target Boltzmann-Gibbs measures

$$\mu_n(dx) \propto 1_{A_n}(x)\ \lambda(dx)$$

associated with a sequence of decreasing subsets $A_n \downarrow$, and some reference measure λ, then we find that $\mu_n = \eta_n$ and $\mathcal{Z}_n = \lambda(A_n)$, as soon as the potential functions in (1.4) and (2.1) are chosen, so that

$$G_n = h_n = 1_{A_{n+1}}.$$

This stochastic model arises in several application domains. In computer science literature, the corresponding particle approximation models are sometimes called subset methods, sequential sampling plans, randomized algorithms, or level splitting algorithms. They were used to solve complex NP-hard combinatorial counting problems [55], extreme quantile probabilities [9, 60], and uncertainty propagations in numerical codes [7].

2.2.3 Stochastic Optimization

If we choose Markov chain Monte Carlo type local moves $\mu_n = \mu_n M_n$ with some prescribed target Boltzmann-Gibbs measures

$$\mu_n(dx) \propto e^{-\beta_n V(x)}\ \lambda(dx)$$

associated with a sequence of increasing inverse temperature parameters $\beta_n \uparrow$, and some reference measure λ, then we find that $\mu_n = \eta_n$ and

$\mathcal{Z}_n = \lambda(e^{-\beta_n V})$ as soon as the potential functions in (1.4) and (2.1) are chosen, so that

$$G_n = h_n = e^{-(\beta_{n+1}-\beta_n)V}.$$

For instance, we can assume that the Markov transition $M_n = \mathcal{M}_{n,\beta_n}^{m_n}$ is the m_n-iterate of the following Metropolis-Hasting transitions

$$\begin{aligned}\mathcal{M}_{n,\beta_n}(x,dy) \;=\; & K_n(x,dy)\min\left(1,e^{-\beta_n(V(y)-V(x))}\right)\\ & +\left(1-\int_z K_n(x,dz)\;\min\left(1,e^{-\beta_n(V(z)-V(x))}\right)\right)\delta_x(dy).\end{aligned}$$

We finish with an assorted collection of enriching comments on interacting Markov chain Monte Carlo algorithms associated with the Feynman-Kac models described above.

Conventional Markov chain Monte Carlo methods with time varying target measures μ_n can be seen as a single particle model with only mutation explorations according to the Markov transitions $M_n = K_n^{m_n}$, where $K_n^{m_n}$ stands for the iteration of an MCMC transition K_n s.t. $\mu_n = \mu_n K_n$. In this situation, we choose a judicious increasing sequence m_n so that the non homogeneous Markov chain is sufficiently stable, even if the target measures become more and more complex to sample. When the target measure is fixed, say of the form μ_T for some large T, the MCMC sampler again uses a single particle, which behaves as a Markov chain with time homogeneous transitions M_T. The obvious drawback with these two conventional MCMC samplers is that the user does not know how many steps are really needed to be close to the equilibrium target measure. A wrong choice will return samples with a distribution far from the desired target measure.

Interacting MCMC methods run a population of MCMC samplers that interact with each other through a recycling-updating mechanism so that the occupation measure of the current measure converges to the target measure, when we increase the population sizes. In contrast with conventional MCMC methods, there are no burn-in time questions, nor any quantitative analysis to estimate the convergence to equilibrium of the MCMC chain.

2.2.4 Island Particle Models

Here, we provide a brief discussion on interacting colonies and island particle models arising in mathematical biology and evolutionary computing literature [87, 96]. The evolution of these stochastic island models is again defined in terms of a free evolution and a selection transition. During the free evolution, each island evolves separately as a single mean field particle model with mutation-selection mechanism between the individual and the island population. The selection pressure between islands is related to the average fitness of the individuals in the island population. A colony with poor fitness is killed, and replaced by brand new generations of "improved" individuals coming from better fitted islands.

For any measurable function f_n on E_n, we set

$$X_n^{(0)} = X_n \in E_n^{(0)} := E_n \quad \text{and} \quad f_n^{(0)} = f_n$$

and we denote by

$$X_n^{(1)} = (X_n^{(1,i)})_{1 \leq i \leq N_1} \in E_n^{(1)} := (E_n^{(0)})^{N_1}$$

the N_1-particle model associated with the reference Markov chain $X_n^{(0)}$, and the potential function $G_n^{(0)}$.

To go one step further, we denote by $f_n^{(1)}$ the empirical mean valued function on $E_n^{(1)}$ defined by

$$f_n^{(1)}(X_n^{(1)}) = \frac{1}{N_1} \sum_{i=1}^{N_1} f_n^{(0)}(X_n^{(1,i)}).$$

In this notation, the potential value of the random state $X_n^{(1)}$ is given by the formula

$$G_n^{(1)}(X_n^{(1)}) := \frac{1}{N_1} \sum_{i=1}^{N} G_n^{(0)}(X_n^{(1,i)}).$$

By construction, we have the almost sure property

$$N_1 = 1 \Longrightarrow X_n^{(0)} = X_n^{(1)} \quad \text{and} \quad G_n^{(1)}(X_n^{(1)}) = G_n^{(0)}(X_n^{(0)}).$$

More interestingly, by the unbiased properties (1.16) we have *for any population size* N_1

$$\mathbb{E}\left(f_n^{(1)}(X_n^{(1)})\prod_{0\leq p<n}G_p^{(1)}(X_p^{(1)})\right)=\mathbb{E}\left(f_n^{(0)}(X_n^{(0)})\prod_{0\leq p<n}G_p^{(0)}(X_p^{(0)})\right).$$

Iterating this construction, we let

$$X_n^{(2)}=(X_n^{(2,i)})_{1\leq i\leq N_2}\in E_n^{(2)}:=(E_n^{(1)})^{N_2}$$

represent the N_2-particle model associated with the reference Markov chain $X_n^{(1)}$, and the potential function $G_n^{(1)}$. For any function $f_n^{(1)}$ on $E_n^{(1)}$, we denote by $f_n^{(2)}$ the empirical mean valued function on $E_n^{(2)}$ defined by

$$f_n^{(2)}(X_n^{(2)})=\frac{1}{N_2}\sum_{i=1}^{N_2}f_n^{(1)}(X_n^{(2,i)}).$$

In this notation, the potential value of the random state $X_n^{(2)}$ is given by the formula

$$G_n^{(2)}(X_n^{(2)}):=\frac{1}{N_2}\sum_{i=1}^{N}G_n^{(1)}(X_n^{(2,i)})$$

and *for any population size* N_2

$$\mathbb{E}\left(f_n^{(2)}(X_n^{(2)})\prod_{0\leq p<n}G_p^{(2)}(X_p^{(2)})\right)=\mathbb{E}\left(f_n^{(0)}(X_n^{(0)})\prod_{0\leq p<n}G_p^{(0)}(X_p^{(0)})\right).$$

2.2.5 Particle Markov Chain Monte Carlo Methods

Now, we present an interacting particle version of the particle Markov chain Monte Carlo method developed in the recent seminal article by Andrieu, Doucet, and Holenstein [2].

We consider a collection of Markov transition and positive potential functions $(M_{\theta,n},G_{\theta,n})$ that depend on some random variable $\Theta=\theta$, with distribution ν on some state space S. We let $\eta_{\theta,n}$ be the n-time marginal of the Feynman-Kac measures defined as in (1.4), by replacing (M_n,G_n) by $(M_{\theta,n},G_{\theta,n})$. We also consider the probability distribution

$P(\theta, d\xi)$ of the N-particle model

$$\xi := (\xi_{\theta,0}, \xi_{\theta,1}, \ldots, \xi_{\theta,T})$$

on the interval $[0,T]$, with mutation transitions $M_{\theta,n}$, and potential selection functions $G_{\theta,n}$, with $n \leq T$. We fix a time horizon T, and for any $0 \leq n \leq T$, we set

$$\mu_n(d(\xi,\theta)) = \frac{1}{\mathcal{Z}_n} \left\{ \prod_{0 \leq p < n} h_p(\xi,\theta) \right\} \lambda(d(\xi,\theta)) \tag{2.2}$$

with some sequence of bounded nonnegative potential functions

$$h_n : (\xi,\theta) \in \left(\prod_{0 \leq p \leq T} E_p^N \right) \times S \mapsto h_n(\xi,\theta) \in (0,\infty)$$

and with the reference measure λ given by

$$\lambda(d(\xi,\theta)) = \nu(d\theta) P(\theta, d\xi)$$

and some normalizing constants $\mathcal{Z}_n$. Firstly, we observe that these target measures have the same form as in (2.1). Thus, they can be sampled using the Interacting Markov chain Monte Carlo methodology presented in subsection 2.2.1.

Now, we examine the situation where h_p is given by the empirical mean value of the potential function $G_{\theta,p}$ w.r.t. the occupation measures $\eta_{\theta,p}^N$ of the N-particle model $\xi_{\theta,p} = (\xi_{\theta,p}^i)_{1 \leq i \leq N}$ associated with the realization $\Theta = \theta$; more formally, we have that

$$h_p(\xi,\theta) = \frac{1}{N} \sum_{1 \leq i \leq N} G_{\theta,p}(\xi_{\theta,p}^i) = \eta_{\theta,p}^N(G_{\theta,p}).$$

Using the unbiased property of the particle free energy models presented in (1.16), we clearly have

$$\begin{aligned} \int P(\theta, d\xi) \left\{ \prod_{0 \leq p < n} h_p(\xi,\theta) \right\} &= \mathbb{E} \left(\prod_{0 \leq p < n} \eta_{\theta,p}^N(G_{\theta,p}) \right) \\ &= \prod_{0 \leq p < n} \eta_{\theta,p}(G_{\theta,p}) \end{aligned}$$

from which we conclude that the Θ-marginal of μ_n is given by the following equation

$$(\mu_n \circ \Theta^{-1})(d\theta) = \frac{1}{\mathcal{Z}_n} \left\{ \prod_{0 \leq p < n} \eta_{\theta,p}(G_{\theta,p}) \right\} \nu(d\theta).$$

We end with some comments on these distributions.

As the initiated reader may certainly have noticed, the marginal analysis derived above coincides with one developed in subsection 2.2.4 dedicated to island particle models.

The measures μ_n introduced in (2.2) can be approximated using the interacting Markov chain Monte Carlo methodology presented in subsection 2.2.1, or the particle MCMC methods introduced in the article [2].

Last but not least, we observe that

$$\prod_{0 \leq p < n} \eta_{\theta,p}(G_{\theta,p}) = \mathbb{E}\left(\prod_{0 \leq p < n} G_{\theta,p}(X_{\theta,p}) \right) = \mathcal{Z}_n(\theta)$$

where $X_{\theta,n}$ stands for the Markov chain with transitions $M_{\theta,n}$, and initial distribution $\eta_{\theta,0}$. In the right-hand side (r.h.s.) of the above displayed formulae, $\mathcal{Z}_n(\theta)$ stands for the normalizing constant of the Feynman-Kac measures defined as in (1.4), by replacing (M_n, G_n) by $(M_{\theta,n}, G_{\theta,n})$. This shows that

$$(\mu_n \circ \Theta^{-1})(d\theta) = \frac{1}{\mathcal{Z}_n} \; \mathcal{Z}_n(\theta) \; \nu(d\theta).$$

The goal of some stochastic optimization problems is to extract the parameter θ that minimizes some mean value functional of the form

$$\theta \mapsto \mathcal{Z}_n(\theta) = \mathbb{E}\left(\prod_{0 \leq p < n} G_{\theta,n}(X_{\theta,p}) \right).$$

For convex functionals, we can use gradient type techniques using the Backward Feynman-Kac derivative interpretation models developed in subsection 2.4.1 (see also [23, 24, 25]).

When ν is the uniform measure over some compact set S, an alternative approach is to estimate the measures (2.2) by some empirical measure

$$\frac{1}{N}\sum_{1\leq i\leq N}\delta_{(\xi_n^{(i)},\theta_n^{(i)})}\in\mathcal{P}\left(\left(\prod_{0\leq p\leq n}E_p^N\right)\times S\right)$$

and to select the sampled state

$$\begin{aligned}(\xi_n^{(i)},\theta_n^{(i)}) &:= (((\xi_{0,n}^{(i,j)})_{1\leq j\leq N},(\xi_{1,n}^{(i,j)})_{1\leq j\leq N},\ldots,(\xi_{n,n}^{(i,j)})_{1\leq j\leq N}),\theta_n^{(i)})\\ &\in ((E_0^N\times E_1^N\times\cdots\times E_n^N)\times S)\end{aligned}$$

that maximizes the empirical objective functional

$$i\in\{1,\ldots,N\}\mapsto\prod_{0\leq p<n}\frac{1}{N}\sum_{1\leq j\leq N}G_{\theta_n^{(i)},p}(\xi_{p,n}^{(i,j)},\theta_n^{(i)}).$$

2.2.6 Markov Bridges and Chains with Fixed Terminal Value

In many applications, it is important to sample paths of Markov chains with prescribed fixed terminal conditions.

When the left end starting point is distributed w.r.t. to a given regular probability measure π, we can use the time reversal Feynman-Kac formula presented in [18]. More precisely, for time homogeneous models $(G_n,M_n)=(G,M)$ in transition spaces, if we consider the Metropolis-Hasting ratio

$$G(x_1,x_2)=\frac{\pi(dx_2)K(x_2,dx_1)}{\pi(dx_1)M(x_1,dx_2)}$$

then we find that

$$\mathbb{Q}_n=\mathrm{Law}_\pi^K((X_0,\ldots,X_n)\mid X_n=x_n)$$

where Law_π^K stands for the distribution of the Markov chain starting with an initial condition π and evolving according to some Markov transition K. The proofs of these formulae are rather technical, so we refer the reader to the article [18] and to the book [17].

For initial and terminal fixed end-points, we need to consider the paths distribution of Markov bridges. As we mentioned in the introduction, these Markov bridges are particular instances of the reference Markov chains of the abstract Feynman-Kac model (1.4). Depending on the choice of the potential functions in (1.4), these Markov bridge models can be associated with several application domains, including filtering problems or rare event analysis of bridge processes.

We assume that the elementary Markov transitions M_n of the chain X_n satisfy the regularity condition (1.10) for some density functions H_n and some reference measure λ_n. In this situation, the semigroup Markov transitions $M_{p,n+1} = M_{p+1}M_{p+2}\ldots M_{n+1}$ are absolutely continuous with respect to the measure λ_{n+1}, for any $0 \leq p \leq n$, and we have

$$M_{p,n+1}(x_p, dx_{n+1}) = H_{p,n+1}(x_p, x_{n+1})\ \lambda_{n+1}(dx_{n+1})$$

with the density function

$$H_{p,n+1}(x_p, x_{n+1}) = M_{p,n}(H_{n+1}(.,x_{n+1}))(x_p).$$

Thanks to these regularity conditions, we readily check that the paths distribution of a Markov bridge starting at x_0 and ending at x_{n+1} at the final time horizon $(n+1)$ are given by

$$\begin{aligned}
&\mathbb{B}_{(0,x_0),(n+1,x_{n+1})}(d(x_1,\ldots,x_n)) \\
&\quad := \mathbb{P}((X_1,\ldots,X_n) \in d(x_1,\ldots,x_n) \mid X_0 = x_0,\ X_{n+1} = x_{n+1}) \\
&\quad = \prod_{1\leq p\leq n} M_p(x_{p-1}, dx_p)\ \frac{dM_{p,n+1}(x_p,.)}{dM_{p-1,n+1}(x_{p-1},.)}(x_{n+1}) \\
&\quad = \prod_{1\leq p\leq n} \frac{M_p(x_{p-1}, dx_p)\ H_{p,n+1}(x_p, x_{n+1})}{M_p(H_{p,n+1}(.,x_{n+1}))(x_{p-1})}.
\end{aligned}$$

Using some abuse of Bayesian notation, we can rewrite these formulae as follows

$$\begin{aligned}
&p((x_1,\ldots,x_n) \mid (x_0,x_{n+1})) \\
&\quad = \frac{p(x_{n+1}|x_n)}{p(x_{n+1}|x_{n-1})} p(x_n|x_{n-1}) \ldots \frac{p(x_{n+1}|x_p)}{p(x_{n+1}|x_{p-1})} p(x_p|x_{p-1}) \\
&\qquad \ldots \frac{p(x_{n+1}|x_1)}{p(x_{n+1}|x_0)} p(x_1|x_0)
\end{aligned}$$

with

$$\frac{dM_p(x_{p-1},.)}{d\lambda_p}(x_p) = p(x_p|x_{p-1}) \quad \text{and} \quad H_{p,n+1}(x_p,x_{n+1}) = p(x_{n+1}|x_p).$$

For linear-Gaussian models, the Markov bridge transitions

$$\frac{M_p(x_{p-1},dx_p)\ H_{p,n+1}(x_p,x_{n+1})}{M_p(H_{p,n+1}(.,x_{n+1}))(x_{p-1})} = \frac{p(x_{n+1}|x_p)}{p(x_{n+1}|x_{p-1})} p(x_p|x_{p-1})\lambda_p(dx_p)$$

can be explicitly computed using the traditional regression formula, or, equivalently, the updating step of the Kalman filter.

2.3 Rare Event Analysis

2.3.1 Importance Sampling and Twisted Measures

Computing the probability of some events of the form $\{V_n(X_n) \geq a\}$, for some energy-like function V_n and some threshold a is often performed using the importance sampling distribution of the state variable X_n with some multiplicative Boltzmann weight function $e^{\beta V_n(X_n)}$ associated with some temperature parameter β. These twisted measures can be described by a Feynman-Kac model in transition space by setting

$$G_n(X_{n-1},X_n) = e^{\beta[V_n(X_n)-V_{n-1}(X_{n-1})]}.$$

For instance, it is easily checked that

$$\begin{aligned}\mathbb{P}(V_n(X_n) \geq a) &= \mathbb{E}(f_n(X_n)\ e^{V_n(X_n)}) \\ &= \mathbb{E}\left(\mathbf{f_n}(\mathbf{X}_n) \prod_{0\leq p<n} G_p(\mathbf{X}_p)\right)\end{aligned}$$

with

$$\mathbf{X}_n = (X_n, X_{n+1}) \quad \text{and} \quad G_n(\mathbf{X}_n) = e^{V_{n+1}(X_{n+1})-V_n(X_n)}$$

and the test function

$$f_n(\mathbf{X}_n) = 1_{V_n(X_n)\geq a}\ e^{-V_n(X_n)}.$$

In the same vein, we have

$$\begin{aligned}
&\mathbb{E}(\varphi_n(X_0,\ldots,X_n) \mid V_n(X_n) \geq a) \\
&\quad = \mathbb{E}(F_{n,\varphi_n}(X_0,\ldots,X_n)\ e^{V_n(X_n)})/\mathbb{E}(F_{n,1}(X_0,\ldots,X_n)\ e^{V_n(X_n)}) \\
&\quad = \mathbb{Q}_n(F_{n,\varphi_n})/\mathbb{Q}_n(F_{n,1})
\end{aligned}$$

with the function

$$F_{n,\varphi_n}(X_0,\ldots,X_n) = \varphi_n(X_0,\ldots,X_n)1_{V_n(X_n)\geq a}e^{-V_n(X_n)}.$$

We illustrate these rather abstract formulae with a Feynman-Kac formulation of European style call options with exercise price a at time n. The prices of these financial contracts are given by formulae of the following form

$$\begin{aligned}
&\mathbb{E}((V_n(X_n) - a)_+) \\
&\quad = \mathbb{E}((V_n(X_n) - a)\ 1_{V_n(X_n)\geq a}) \\
&\quad = \mathbb{P}(V_n(X_n) \geq a) \times \mathbb{E}((V_n(X_n) - a) \mid V_n(X_n) \geq a).
\end{aligned}$$

It is now a simple exercise to check that these formulae fit with the Feynman-Kac importance sampling model discussed above. Further details on these models, including applications in fiber optics communication and financial risk analysis can also be found in [6, 26, 27].

2.3.2 Rare Event Excursion Models

If we consider Markov excursion type models between a sequence of decreasing subsets A_n or hitting an absorbing level B, then choosing an indicator potential function that detects if the n-th excursion hits A_n before B we find that

$$\mathbb{Q}_n = \mathrm{Law}(X \text{ hits } A_n \mid X \text{ hits } A_n \text{ before } B)$$

and

$$\mathbb{P}(X \text{ hits } A_n \text{ before } B) = \mathbb{E}\left(\prod_{0\leq p\leq n} 1_{A_p}(X_{T_p})\right) = \mathbb{E}\left(\prod_{0\leq p<n} G_p(\mathbf{X}_p)\right)$$

with the random times

$$T_n := \inf\{p \geq T_{n-1} : X_p \in (A_n \cup B)\}$$

and the excursion models

$$\mathbf{X}_n = (X_p)_{p\in[T_n,T_{n+1}]} \quad \& \quad G_n(\mathbf{X}_n) = 1_{A_{n+1}}(X_{T_{n+1}}).$$

In this notation, it is also easily checked that

$$\mathbb{E}(\mathbf{f_n}(X_{[0,T_{n+1}]}) \mid X \text{ hits } A_n \text{ before } B) = \mathbb{Q}_n(\mathbf{f}_n).$$

For a more thorough discussion on these excursion particle models, we refer the reader to the series of articles [8, 10, 17, 37, 64].

2.4 Sensitivity Measures

2.4.1 Kinetic Sensitivity Measures

We let $\theta \in \mathbb{R}^d$ be some parameter that may represent some kinetic type parameters related to the free evolution model or to the adaptive potential functions. We assume that the free evolution model $X_k^{(\theta)}$ associated to some value of the parameter θ, is given by a one-step probability transition of the form

$$M_k^{(\theta)}(x, dx') := \mathbb{P}(X_k^{(\theta)} \in dx' | X_{k-1}^{(\theta)} = x) = H_k^{(\theta)}(x, x')\ \lambda_k(dx')$$

for some positive density functions $H_k^{(\theta)}$ and some reference measures λ_k. We also consider a collection of functions $G_k^{(\theta)} = e^{-V_k^{(\theta)}}$ that depend on θ. We also assume that the gradient and the Hessian of the logarithms of these functions w.r.t. the parameter θ are well defined. We let Γ_n^θ be the Feynman-Kac measure associated with a given value of θ defined for any function $\mathbf{f_n}$ on the path space $\mathbf{E_n}$ by

$$\Gamma_n^\theta(\mathbf{f_n}) = \mathbb{E}\left(\mathbf{f_n}(X_0^{(\theta)}, \ldots, X_n^{(\theta)}) \prod_{0\leq p<n} G_p^{(\theta)}(X_p^{(\theta)})\right). \tag{2.3}$$

We denote by $\Gamma_n^{(\theta,N)}$ the N-particle approximation measures associated with a given value of the parameter θ and defined in (1.18).

By using simple derivation calculations, we prove that the first-order derivative of the option value w.r.t. θ is given by

$$\begin{aligned}\nabla\Gamma_n^{(\theta)}(\mathbf{f_n}) &= \Gamma_n^{(\theta)}(\mathbf{f_n}\Lambda_n^{(\theta)})\\ \nabla^2\Gamma_n^{(\theta)}(\mathbf{f_n}) &= \Gamma_n^{(\theta)}[\mathbf{f_n}(\nabla\mathbb{L}_n^{(\theta)})'(\nabla\mathbb{L}_n^{(\theta)}) + \mathbf{f_n}\nabla^2\mathbb{L}_n^{(\theta)}]\end{aligned}$$

with

$$\Lambda_n^{(\theta)} := \nabla\mathbb{L}_n^{(\theta)}$$

and the additive functional

$$\mathbb{L}_n^{(\theta)}(x_0,\ldots,x_n) := \sum_{p=1}^{n}\log\left(G_{p-1}^{(\theta)}(x_{p-1})H_p^{(\theta)}(x_{p-1},x_p)\right).$$

These quantities are approximated by the unbiased particle models

$$\begin{aligned}\nabla_N\Gamma_n^{(\theta)}(\mathbf{f_n}) &:= \Gamma_n^{(\theta,N)}(\mathbf{f_n}\Lambda_n^{(\theta)})\\ \nabla_N^2\Gamma_n^{(\theta)}(\mathbf{f_n}) &= \Gamma_n^{(\theta,N)}[\mathbf{f_n}(\nabla\mathbb{L}_n^{(\theta)})'(\nabla\mathbb{L}_n^{(\theta)}) + \mathbf{f_n}\nabla^2\mathbb{L}_n^{(\theta)}].\end{aligned}$$

For a more thorough discussion on these Backward Feynman-Kac models, we refer the reader to [23, 24, 25].

2.4.2 Gradient Estimation of a Markov Semigroup

We assume that the underlying stochastic evolution is given by an iterated $\mathbb{R}^d$-valued random process given by the following equation

$$X_{n+1} := F_n(X_n) = (F_n \circ F_{n-1} \circ \cdots \circ F_0)(X_0) \tag{2.4}$$

starting at some random state X_0, with a sequence of random smooth functions of the form

$$F_n(x) = \mathcal{F}_n(x, W_n) \tag{2.5}$$

with some smooth collection of functions

$$\mathcal{F}_n : (x,w) \in \mathbb{R}^{d+d'} \mapsto \mathcal{F}_n(x,w) \in \mathbb{R}^d$$

and some collection of independent, and independent of s, random variables W_n taking values in some $\mathbb{R}^{d'}$, with $d' \geq 1$. The semigroup of the

Markov chain X_n is the expectation operator defined for any regular function f_n and any state x by

$$P_{n+1}(f_{n+1})(x) := \mathbb{E}(f_{n+1}(X_{n+1}) \mid X_0 = x) = \mathbb{E}(f(X_{n+1}(x)))$$

with the random flows $(X_n(x))_{n\geq 0}$ defined for any $n \geq 0$ by the following equation

$$X_{n+1}(x) = F_n(X_n(x))$$

with the initial condition $X_0(x) = x$.

By construction, for any $1 \leq i, j \leq d$ and any $x \in \mathbb{R}^d$ we have the first variational equation

$$\frac{\partial X_{n+1}^i}{\partial x^j}(x) = \sum_{1\leq k\leq d} \frac{\partial F_n^i}{\partial x^k}(X_n(x)) \frac{\partial X_n^k}{\partial x^j}(x). \tag{2.6}$$

This clearly implies that

$$\frac{\partial P_{n+1}(f)}{\partial x^j}(x) = \mathbb{E}\left(\sum_{1\leq i\leq d} \frac{\partial f}{\partial x^i}(X_{n+1}(x)) \ \frac{\partial X_{n+1}^i}{\partial x^j}(x) \right). \tag{2.7}$$

We denote by $V_n = (V_n^{(i,j)})_{1\leq i,j\leq d}$ and $A_n = (A_n^{(i,j)})_{1\leq i,j\leq d}$ the random $(d \times d)$ matrices with the i-th line and j-th column entries

$$V_n^{(i,j)}(x) = \frac{\partial S_n^i}{\partial x^j}(x)$$

and

$$A_n^{(i,j)}(x) = \frac{\partial F_n^i}{\partial x^j}(x) = \frac{\partial \mathcal{F}_n^i(.,W_n)}{\partial x^j}(x) := \mathcal{A}_n^{(i,j)}(x, W_n).$$

In this notation, the Equation (2.6) can be rewritten in terms of the following random matrix formulae

$$V_{n+1}(x) \ = \ A_n(X_n(x)) \ V_n(x) := \prod_{p=0}^{n} A_p(X_p(x)) \tag{2.8}$$

with a product $\prod_{p=0}^n A_p$ of noncommutative random elements A_p taken in the order A_n, A_{n-1},..., A_0. In the same way, the Equation (2.7) can be rewritten as

$$\nabla P_{n+1}(f_{n+1})(x) = \mathbb{E}\left(\nabla f_{n+1}(X_{n+1}) \ V_{n+1} \mid X_0 = x\right) \tag{2.9}$$

with

$$V_{n+1} := \prod_{0 \leq p \leq n} A_p(X_p).$$

We equip the space $\mathbb{R}^d$ with some norm $\|.\|$. We assume that for any state U_0 in the unit sphere $\mathcal{S}^{d-1}$, we have

$$\|V_{n+1}U_0\| > 0.$$

In this situation, we have the multiplicative formulae

$$\nabla f_{n+1}(X_{n+1})\ V_{n+1}U_0 = [\nabla f_{n+1}(X_{n+1})U_{n+1}] \prod_{0 \leq p \leq n} \|A_p(X_p)U_p\|$$

with the well defined $\mathcal{S}^{d-1}$-valued Markov chain defined by

$$U_{n+1} = A_n(X_n)U_n/\|A_n(X_n)U_n\| \left(\Leftrightarrow U_{n+1} = \frac{V_{n+1}\ U_0}{\|V_{n+1}U_0\|} \right).$$

If we choose $U_0 = u_0$, then we obtain the following Feynman-Kac interpretation of the gradient of a semigroup

$$\nabla P_{n+1}(f_{n+1})(x)u_0 = \mathbb{E}\left(F_{n+1}(\mathcal{X}_{n+1}) \prod_{0 \leq p \leq n} \mathcal{G}_p(\mathcal{X}_p) \right). \tag{2.10}$$

In the above display, $\mathcal{X}_n$ is the Markov chain sequence

$$\mathcal{X}_n := (X_n, U_n, W_n)$$

starting at (x, u_0, W_0), and the functions F_{n+1} and G_n are defined by

$$F_{n+1}(x,u,w) := \nabla f_{n+1}(x)u \quad \text{and} \quad \mathcal{G}_n(x,u,w) := \|\mathcal{A}_n(x,w)u\|.$$

In computational physics literature, the mean particle approximations of these non commutative Feynman-Kac models are often referred to as Resampled Monte Carlo methods [94].

Roughly speaking, in addition to the fact that formula (2.10) provides an explicit functional Feynman-Kac description of the the gradient of a Markov semigroup, the random evolution model U_n on the unite sphere may be degenerate. More precisely, the Markov chain $\mathcal{X}_n = (X_n, U_n, W_n)$ may not satisfy the regularity properties stated in

subsection 3.4.1. We end with some rather crude upper bounds that can be estimated uniformly w.r.t. the time parameter under appropriate regularity conditions on the reduced Markov chain model (X_n, W_n). First, we notice that

$$\mathcal{G}_n(x,u,w) := \|\mathcal{A}_n(x,w)u\| \leq G_n(x,w) := \|\mathcal{A}_n(x,w)\| \\ := \sup_{u\in\mathcal{S}^{d-1}} \|\mathcal{A}_n(x,w)\ u\|.$$

This implies that

$$\|\nabla P_{n+1}(f_{n+1})(x)\| := \sup_{1\leq i\leq d}\left|\frac{\partial}{\partial x^i}\ P_{n+1}(f_{n+1})(x)\right| \\ \leq\ \|F_{n+1}\| \times \mathbb{E}\left(\prod_{0\leq p\leq n} G_p(X_p,W_p)\right).$$

The r.h.s. functional expectation in the above equation can be approximated using the particle approximation (1.15) of the multiplicative Feyman-Kac formulae (1.9), with reference Markov chain (X_n, W_n) and potential functions G_n.

2.5 Partial Observation Models

2.5.1 Nonlinear Filtering Models

Next, we introduce one of the most important examples of an estimation problem with partial observation, namely the nonlinear filtering model. This model has been the starting point of the application of particle models to engineering sciences, and more particularly to advanced signal processing.

The first rigorous investigation of stochastic modeling, and the rigorous theoretical analysis of particle filters began in the mid-1990's in the article [15]. For a detailed discussion on the application domains of particle filtering, with a precise bibliography we refer the reader to any of the following references [17, 20, 38, 44].

The typical model is given by a reference Markov chain model X_n, and some partial and noisy observation Y_n. The pair process (X_n, Y_n) usually forms a Markov chain on some product space $E^X \times E^Y$ with

elementary transitions given as

$$\mathbb{P}((X_n, Y_n) \in d(x,y) \mid (X_{n-1}, Y_{n-1})) = M_n(X_{n-1}, dx) g_n(x,y) \lambda_n(dy) \tag{2.11}$$

for some positive likelihood function g_n, and some reference probability measure λ_n on E^Y, and the elementary Markov transitions M_n of the Markov chain X_n. If we take

$$G_n(x_n) = p_n(y_n|x_n) = g_n(x_n, y_n) \tag{2.12}$$

the likelihood function of a given observation $Y_n = y_n$ and a signal state $X_n = x_n$ associated with a filtering or an hidden Markov chain problem, then we find that

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

and $\mathcal{Z}_{n+1}$ is the density of the sequence $y_0, \ldots, y_n$ w.r.t. the product measure $\lambda_0 \otimes \cdots \otimes \lambda_n$ evaluated on the sequence of observation $Y_p = y_p,\ p \leq n$.

In this context, the optimal one-step predictor η_n and the optimal filter $\widehat{\eta}_n$ are given by the n-th time marginal distribution

$$\eta_n^{[y_0,\ldots,y_{n-1}]} = \eta_n = \text{Law}\,(X_n \mid \forall 0 \leq p < n \, Y_p = y_p) \tag{2.13}$$

and

$$\widehat{\eta}_n^{[y_0,\ldots,y_n]} = \widehat{\eta}_n = \Psi_{G_n}(\eta_n) = \text{Law}\,(X_n \mid \forall 0 \leq p \leq n \quad Y_p = y_p)\,. \tag{2.14}$$

In Bayesian literature, the potential function $G_n(x_n) = g_n(x_n, y_n)$ is often written as $p_n(y_n|x_n)$. In this slight abuse of notation, the normalizing constant $\mathcal{Z}_{n+1}$ is also written as $\mathcal{Z}_{n+1} = p_n(y_0, \ldots, y_n)$.

Remark 2.1. We can combine these filtering models with the probability restriction models discussed in subsection 2.2.2, or with the rare event analysis presented in subsection 2.3. For instance, if we replace the potential likelihood function G_n defined in (2.12) by the function

$$G_n(x_n) = g_n(x_n, y_n)\ 1_{A_n}(x_n)$$

then we find that

$$\mathbb{Q}_n = \text{Law}((X_0, \ldots, X_n) \mid \forall 0 \leq p < n \quad Y_p = y_p,\ X_p \in A_p).$$

2.5.2 Approximated Filtering Models

We return to the stochastic filtering model discussed in subsection 2.5.1. In some instances, the likelihood functions $x_n \mapsto g_n(x_n, y_n)$ in (2.12) are computationally intractable, or too expensive to evaluate.

To solve this problem, a natural solution is to sample pseudo-observations. The central idea is to sample the signal-observation Markov chain

$$\mathbf{X}_n = (X_n, Y_n) \in \mathbf{E}^{\mathbf{X}} = (E^X \times E^Y)$$

and compare the values of the sampled observations with the real observations.

To describe these models with some precision, we notice that the transitions of $\mathbf{X}_n$ are given by

$$\mathbf{M_n}(\mathbf{X_{n-1}}, d(x,y)) = M_n(X_{n-1}, dx) g_n(x,y) \lambda_n(dy).$$

To simplify the presentation, we further assume that $E^Y = \mathbb{R}^d$, for some $d \geq 1$, and we let g be a Borel bounded non negative function such that

$$\int g(u)du = 1 \quad \int ug(u)du = 0 \quad \text{and} \quad \int |u|^3 g(u)du < \infty.$$

Then, we set for any $\epsilon > 0$, and any $\mathbf{x} = (x,y) \in (E^X \times E^Y)$

$$g_{\epsilon,n}((x,y),z) = \epsilon^{-d}\, g\left((y-z)/\epsilon\right).$$

Finally, we let $(\mathbf{X_n}, \mathbf{Y_n^\epsilon})$ be the Markov chain on the augmented state space $(\mathbf{E^X} \times E^Y) = ((E^X \times E^Y) \times E^Y)$ with transitions given

$$\begin{aligned} &\mathbb{P}((\mathbf{X_n}, \mathbf{Y_n^\epsilon}) \in d(\mathbf{x}, y) \mid (\mathbf{X_{n-1}}, \mathbf{Y_{n-1}^\epsilon})) \\ &\qquad = \mathbf{M_n}(\mathbf{X_{n-1}}, \mathbf{dx}) g_{\epsilon,n}(\mathbf{x}, y) dy. \end{aligned} \tag{2.15}$$

This approximated filtering problem has exactly the same form as the one introduced in (2.11). Here, the particle approximation models are defined in terms of signal-observation valued particles, and the selection potential function is given by the pseudo-likelihood functions $g_{\epsilon,n}(\cdot, y_n)$, where y_n stands for the value of the observation sequence at time n.

For a more detailed discussion on these particle models, including the convergence analysis of the approximated filtering model, we refer

the reader to the articles [33, 34]. These particle models are sometimes called convolution particle filters [95]. In Bayesian literature, these approximated filtering models are termed Approximate Bayesian Computation (and often abbreviated with *the acronym ABC*).

2.5.3 Parameter Estimation in Hidden Markov Chain Models

We consider a pair signal-observation filtering model (X,Y) that depends on some random variable Θ with distribution μ on some state space S. Arguing as above, and using the same notations as in subsection 2.5.1, if we take

$$G_{\theta,n}(x_n) = p_n(y_n|x_n,\theta)$$

the likelihood function of a given observation $Y_n = y_n$ and a signal state $X_n = x_n$ and a realization of the parameter $\Theta = \theta$, then the n-th time marginal of $\mathbb{Q}_n$ is given by

$$\eta_{\theta,n} = \mathrm{Law}(X_n \mid \forall 0 \leq p < n \; Y_p = y_p, \theta).$$

Using the multiplicative formula (1.9), we prove that

$$\mathcal{Z}_{n+1}(\theta) = p_n(y_0,\ldots,y_n|\theta) = \prod_{0\leq p\leq n} \eta_{\theta,p}(G_{\theta,p})$$

with

$$\begin{aligned}\eta_{\theta,p}(G_{\theta,p}) &= p(y_p|y_0,\ldots,y_{p-1},\theta)\\ &= \int p(y_p|x_p,\theta)dp(x_p|\theta,y_0,\ldots,y_{p-1})\\ &= \int G_{\theta,p}(x_p)\; \eta_{\theta,p}(dx_p)\end{aligned}$$

from which we conclude that

$$\mathbb{P}(\Theta \in d\theta \mid \forall 0 \leq p \leq n \; Y_p = y_p) = \frac{1}{\mathcal{Z}_n}\mathcal{Z}_n(\theta)\mu(d\theta)$$

with

$$\mathcal{Z}_n := \int \; \mathcal{Z}_n(\theta) \; \mu(d\theta).$$

In some instances, such as in conditionally linear Gaussian models, the normalizing constants $\mathcal{Z}_n(\theta)$ can be computed explicitly, and we can use a Metropolis-Hasting style Markov chain Monte Carlo method to sample the target measures μ_n. As in subsection 2.2.1, we can also turn this scheme into an interacting Markov chain Monte Carlo algorithm.

Indeed, let us choose a Markov chain Monte Carlo type local moves $\mu_n = \mu_n M_n$ with prescribed target measures

$$\mu_n(d\theta) := \frac{1}{\mathcal{Z}_n} \ \mathcal{Z}_n(\theta) \ \mu(d\theta).$$

Notice that

$$\mathcal{Z}_{n+1}(\theta) = \mathcal{Z}_n(\theta) \times \eta_{\theta,n}(G_{\theta,n}) \Rightarrow \mu_{n+1} = \Psi_{G_n}(\mu_n)$$

with the Boltzmann-Gibbs transformations defined in (1.8) associated with the potential function

$$G_n(\theta) = \eta_{\theta,n}(G_{\theta,n}).$$

By construction, we have

$$\mu_{n+1} = \mu_{n+1} M_{n+1} = \Psi_{G_n}(\mu_n) M_{n+1}$$

from which we conclude that

$$\mu_n(f) = \mathbb{E}\left(f(\theta_n) \prod_{0 \leq p < n} G_p(\theta_p) \right) \Big/ \mathbb{E}\left(\prod_{0 \leq p < n} G_p(\theta_p) \right)$$

with the reference Markov chain

$$\mathbb{P}(\theta_n \in d\theta | \theta_{n-1}) = M_n(\theta_{n-1}, d\theta).$$

In addition, we have

$$\mathcal{Z}_{n+1} = \int \mathcal{Z}_n(\theta) \ G_n(\theta) \mu(d\theta) = \mathcal{Z}_n \mu_n(G_n) = \prod_{0 \leq p \leq n} \mu_p(G_p).$$

Remark 2.2. For more general models, we can use the particle Markov chain Monte Carlo methodology presented in subsection 2.2.5. When the likelihood functions are too expensive to evaluate, we can also combine these particle models with the pseudo-likelihood stochastic models (2.15) discussed in subsection 2.5.2.

2.5.4 Interacting Kalman-Bucy Filters

We use the same notation as above, but we assume that $\Theta = (\Theta_n)_{n\geq 0}$ is a random sample of a stochastic process Θ_n taking values in some state spaces S_n. If we consider the Feynman-Kac model associated with the Markov chain $\mathcal{X}_n = (\Theta_n, \eta_{\Theta,n})$ and the potential functions

$$\mathcal{G}_n(\mathcal{X}_n) = \eta_{\Theta,n}(G_{\Theta,n})$$

then we find that

$$\mathbb{Q}_n = \mathrm{Law}(\Theta_0, \ldots, \Theta_n \mid \forall 0 \leq p < n \; Y_p = y_p)$$

and the n-th time marginals are clearly given by

$$\eta_n = \mathrm{Law}(\Theta_n \mid \forall 0 \leq p < n \; Y_p = y_p).$$

Assuming that the pair (X, Y) is a linear and Gaussian filtering model given Θ, the measures $\eta_{\Theta,n}$ coincide with the one-step predictor of the Kalman-Bucy filter, and the potential functions $\mathcal{G}_n(\mathcal{X}_n)$ can be easily computed by Gaussian integral calculations. In this situation, using the same notations as in subsection 2.5.1, the conditional distribution of the parameter Θ is given by a Feynman-Kac model $\mathbb{Q}_n$ of a the free Markov chain $\mathcal{X}_n$ weighted by some Boltzmann-Gibbs exponential weight function

$$\prod_{0\leq p<n} \mathcal{G}_p(\mathcal{X}_p) = p_n(y_0, \ldots, y_n | \Theta_0, \ldots, \Theta_n)$$

that reflects the likelihood of the path sequence $(\Theta_0, \ldots, \Theta_n)$. For a more thorough discussion on these interacting Kalman filters, we refer the reader to Sections 2.6 and 12.6 in the book [17].

2.5.5 Multi-Target Tracking Models

Multiple-target tracking problems deal with correctly estimating several maneuvering and interacting targets simultaneously given a sequence of noisy and partial observations. At every time n, the first moment of the occupation measure $\mathcal{X}_n := \sum_{i=1}^{N_n} \delta_{X_n^i}$ of some spatial branching signal is given for any regular function f by the following

formula:

$$\gamma_n(f) := \mathbb{E}(\mathcal{X}_n(f)) \quad \text{with } \mathcal{X}_n(f) := \int f(x)\ \mathcal{X}_n(dx).$$

For null spontaneous birth measures, these measures coincide with that of an unnormalized Feynman-Kac model with some spatial branching potential functions G_n and some free evolution target model X_n.

In more general situations, the approximate filtering equation is given by the Mahler's multi-objective filtering approximation based on the propagation of the first conditional moments of Poisson approximation models [73, 74]. These evolution equations are rather complex to introduce and are notationally consuming. Nevertheless, as the first moment of evolution of any spatial and marked branching process, they can be abstracted by an unnormalized Feynman-Kac model with nonlinear potential functions [52, 53, 54].

2.5.6 Optimal Stopping Problems with Partial Observations

We consider the partially observed Markov chain model discussed in (2.11). The Snell envelope associated with an optimal stopping problem with finite horizon, payoff style function $f_n(X_n, Y_n)$, and noisy observations Y_n as some Markov process, is given by

$$U_k := \sup_{\tau \in \mathcal{T}_k^Y} \mathbb{E}(f_\tau(X_\tau, Y_\tau)|(Y_0, \ldots, Y_k))$$

where $\mathcal{T}_k^Y$ stands for the set of all stopping times τ taking values in $\{k, \ldots, n\}$, whose values are measurable w.r.t. the sigma field generated by the observation sequence Y_p, from $p = 0$ up to the current time k. We denote by $\eta_n^{[y_0,\ldots,y_{n-1}]}$ and $\widehat{\eta}_n^{[y_0,\ldots,y_n]}$ the conditional distributions defined in (2.13) and (2.14). In this notation, for any $0 \leq k \leq n$ we have that

$$\begin{aligned} &\mathbb{E}(f_\tau(X_\tau, Y_\tau)|(Y_0, \ldots, Y_k)) \\ &\quad = \mathbb{E}(F_\tau(Y_\tau, \widehat{\eta}_\tau^{[Y_0,\ldots,Y_\tau]}) \mid (Y_0, \ldots, Y_k)) \end{aligned} \tag{2.16}$$

with the conditional payoff function

$$F_p(Y_p, \widehat{\eta}_p^{[Y_0,\ldots,Y_p]}) = \int \widehat{\eta}_p^{[Y_0,\ldots,Y_p]}(dx) f_p(X_p, Y_p).$$

It is rather well known that

$$\mathcal{X}_p := (Y_p, \widehat{\eta}_p^{[Y_0,\ldots,Y_p]})$$

is a Markov chain with elementary transitions defined by

$$\begin{aligned}\mathbb{E}[F_p(Y_p, \widehat{\eta}_p^{[Y_0,\ldots,Y_p]}) \mid (Y_{p-1}, \widehat{\eta}_{p-1}^{[Y_0,\ldots,Y_{p-1}]}) = (y, \mu)] \\ = \int \lambda_p(dy_p)\mu M_p(g_p(.,y_p))F_p(y_p, \Psi_{g_p(.,y_p)}(\mu M_p)).\end{aligned}$$

A detailed proof of this assertion can be found in any textbook on advanced stochastic filtering theory. For instance, the book of Runggaldier and Stettner [86] provides a detailed treatment on discrete time non linear filtering, and related partially observed control models.

Roughly speaking, using some abuse of Bayesian notation, we have

$$\begin{aligned}\eta_p^{[y_0,\ldots,y_{p-1}]}(dx_p) &= dp_p(x_p \mid (y_0,\ldots,y_{p-1})) \\ &= \int dp_p(x_p \mid x_{p-1}) \times p_n(x_{p-1} \mid (y_0,\ldots,y_{p-1})) \\ &= \widehat{\eta}_{p-1}^{[y_0,\ldots,y_{p-1}]} M_p(dx_p)\end{aligned}$$

and

$$\begin{aligned}&\Psi_{g_p(.,y_p)}(\widehat{\eta}_{p-1}^{[y_0,\ldots,y_{p-1}]} M_p)(dx_p) \\ &\quad = \frac{p(y_p|x_p)}{\int p_p(y_p \mid x'_p) dp_p(x'_p \mid (y_0,\ldots,y_{p-1}))} dp_p(x_p \mid (y_0,\ldots,y_{p-1})) \\ &\quad = dp_p(x_p \mid (y_0,\ldots,y_{p-1},y_p))\end{aligned}$$

from which we prove that

$$\begin{aligned}\mu M_p(g_p(.,y_p)) &= \int p_p(y_p \mid x_p) dp_p(x_p \mid (y_0,\ldots,y_{p-1})) \\ &= p_p(y_p \mid (y_0,\ldots,y_{p-1}))\end{aligned}$$

and

$$\Psi_{g_p(.,y_p)}(\mu M_p) = \widehat{\eta}_p^{[y_0,\ldots,y_p]}$$

as soon as $\mu = \widehat{\eta}_{p-1}^{[y_0,\ldots,y_{p-1}]} (\Rightarrow \mu M_p = \eta_p^{[y_0,\ldots,y_{p-1}]})$.

From the above discussion, we can rewrite (2.16) as the Snell envelope of a fully observed augmented Markov chain sequence

$$\mathbb{E}(f_\tau(X_\tau, Y_\tau)|(Y_0, \ldots, Y_k)) = \mathbb{E}(F_\tau(\mathcal{X}_\tau) \mid (\mathcal{X}_0, \ldots, \mathcal{X}_k)).$$

The Markov chain $\mathcal{X}_n$ takes values in an infinite dimensional state space, and it can rarely be sampled without some additional level of approximation. Using the N-particle approximation models, we can replace the chain $\mathcal{X}_n$ by the N-particle approximation model defined by

$$\mathcal{X}_n^N := (Y_p, \widehat{\eta}_p^{([Y_0,\ldots,Y_p],N)})$$

where

$$\widehat{\eta}_p^{([Y_0,\ldots,Y_p],N)} := \Psi_{g_p(.,Y_p)}(\eta_{p-1}^{([Y_0,\ldots,Y_{p-1},N)]})$$

stands for the updated measure associated with the likelihood selection functions $g_p(., Y_p)$. The N-particle approximation of the Snell envelope is now given by

$$\mathbb{E}(f_\tau(X_\tau, Y_\tau)|(Y_0, \ldots, Y_k)) \simeq_{N\uparrow\infty} \mathbb{E}(F_\tau(\mathcal{X}_\tau^N) \mid (\mathcal{X}_0^N, \ldots, \mathcal{X}_k^N)).$$

In this interpretation, the N-approximated optimal stopping problem amounts to computing the quantities

$$U_k^N := \sup_{\tau \in \mathcal{T}_k^N} \mathbb{E}(F_\tau(\mathcal{X}_\tau^N) \mid (\mathcal{X}_0^N, \ldots, \mathcal{X}_k^N))$$

where $\mathcal{T}_k^N$ stands for the set of all stopping times τ taking values in $\{k, \ldots, n\}$, whose values are measurable w.r.t. the sigma field generated by the Markov chain sequence $\mathcal{X}_k^N$, from $p = 0$ up to time k.

2.6 Markov Chain Restriction Models

2.6.1 Markov Confinement Models

One of the simplest example of Feynman-Kac conditional distributions is given by choosing indicator functions $G_n = 1_{A_n}$ of measurable subsets $A_n \in \mathcal{E}_n$ s.t. $\mathbb{P}(\forall 0 \leq p < n X_p \in A_p) > 0$. In this situation, it is readily checked that

$$\mathbb{Q}_n = \mathrm{Law}((X_0, \ldots, X_n) \mid \forall 0 \leq p < n \; X_p \in A_p)$$

and

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

This Markov chain restriction model fits into the particle absorption model (1.6) presented in the introduction. For a detailed analysis of these stochastic models, and their particle approximations, we refer the reader to the articles [19, 38, 39], and the book [17].

2.6.2 Directed Polymers and Self-avoiding Walks

The conformation of polymers in a chemical solvent can be seen as the realization of a Feynman-Kac distribution of a free Markov chain weighted by some Boltzmann-Gibbs exponential weight function that reflects the attraction or the repulsion forces between the monomers. For instance, if we consider the historical process

$$\mathbf{X_n} = (X_0, \ldots, X_n)$$

and

$$\mathbf{G_n}(\mathbf{X_n}) = 1_{\notin \{X_p, p<n\}}(X_n)$$

then we find that

$$\begin{aligned} \mathbb{Q}_n &= \mathrm{Law}(\mathbf{X_n} \mid \forall 0 \leq p < n \ \mathbf{X_p} \in A_p) \\ &= \mathrm{Law}((X_0, \ldots, X_n) \mid \forall 0 \leq p < q < n \ X_p \neq X_q) \end{aligned}$$

with the set $A_n = \{\mathbf{G_n} = \mathbf{1}\}$, and the normalizing constants

$$\mathcal{Z}_n = \mathbb{P}(\forall 0 \leq p < q < n \ X_p \neq X_q).$$

2.7 Particle Absorption Models

We return to the particle absorption model (1.6) presented in the introduction. For instance, we can assume that the potential function G_n and Markov transitions M_n are defined by $G_n(x) = e^{-V_n(x)h}$, and

$$M_n(x, dy) = (1 - \lambda_n h)\delta_x(dy) + \lambda_n h K_n(x, dy) \tag{2.17}$$

for some non negative and bounded function V_n, some positive parameter $\lambda_n \leq 1/h$, $h > 0$, and some Markov transition K_n.

The confinement models described above can also be interpreted as a particle absorption model related to hard obstacles. In branching processes and population dynamics literature, the model X_n^c often represents the number of individuals of a given species [51, 57, 88]. Each individual can die or reproduce. The state $0 \in E_n = \mathbb{N}$ is interpreted as a trap, or as a hard obstacle, in the sense that the species disappears as soon as X_n^c hits 0. For a more thorough discussion on particle motions in an absorbing medium with hard and soft obstacles, we refer the reader to the pair of articles [39, 19].

2.7.1 Doob h-Processes

We consider a time homogeneous Feynman-Kac model $(G_n, M_n) = (G, M)$ on some measurable state space E, and we set

$$Q(x, dy) = G(x)M(x, dy).$$

We also assume that G is uniformly bounded above and below by some positive constant, and the Markov transition M is reversible w.r.t. some probability measure μ on E, with $M(x, .) \simeq \mu$ and $dM(x, .)/d\mu \in \mathbb{L}_2(\mu)$. We denote by λ the largest eigenvalue of the integral operator Q on $\mathbb{L}_2$, and by $h(x)$ a positive eigenvector

$$Q(h) = \lambda h.$$

The Doob h-process corresponding to the ground state eigenfunction h defined above is a Markov chain X_n^h with the time homogeneous Markov transition

$$M^h(x, dy) := \frac{1}{\lambda} \times h^{-1}(x)Q(x, dy)h(y) = \frac{M(x, dy)h(y)}{M(h)(x)}$$

and initial distribution $\eta_0^h(dx) \propto h(x)\ \eta_0(dx)$. By construction, we have $G = \lambda h/M(h)$ and therefore

$$\Gamma_n(d(x_0, \ldots, x_n)) = \lambda^n \eta_0(h)\mathbb{P}_n^h(d(x_0, \ldots, x_n))\frac{1}{h(x_n)}$$

where $\mathbb{P}_n^h$ stands for the law of the historical process

$$\mathbf{X_n^h} = (X_0^h, \ldots, X_n^h).$$

We conclude that

$$d\mathbb{Q}_n = \frac{1}{\mathbb{E}(h^{-1}(X_n^h))} h^{-1}(X_n^h) d\mathbb{P}_n^h$$

with the normalizing constants

$$\mathcal{Z}_n = \lambda^n \eta_0(h) \mathbb{E}(h^{-1}(X_n^h)).$$

2.7.2 Yaglom Limits and Quasi-invariant Measures

We return to the time homogeneous Feynman-Kac models introduced in the last subsection. Using the particle absorption interpretation (1.6) we have

$$\mathrm{Law}((X_0^c, \ldots, X_n^c) \mid T^c \geq n) = \frac{1}{\mathbb{E}(h^{-1}(X_n^h))} h^{-1}(X_n^h) d\mathbb{P}_n^h$$

and

$$\mathcal{Z}_n = \mathbb{P}(T^c \geq n) = \lambda^n \eta_0(h) \mathbb{E}(h^{-1}(X_n^h)) \longrightarrow_{n \uparrow \infty} 0. \tag{2.18}$$

Letting $\eta_n^h := \mathrm{Law}(X_n^h)$, we readily prove the following formulae

$$\eta_n = \Psi_{1/h}(\eta_n^h) \quad \text{and} \quad \eta_n^h = \Psi_h(\eta_n).$$

Whenever it exists, the Yaglom limit of the measure η_0 is defined as the limiting of measure

$$\eta_n \longrightarrow_{n \uparrow \infty} \eta_\infty = \Psi_G(\eta_\infty) M \tag{2.19}$$

of the Feynman-Kac flow η_n, when n tends to infinity. We also say that η_0 is a quasi-invariant measure if we have $\eta_0 = \eta_n$, for any time step. When the Feynman-Kac flow η_n is asymptotically stable, in the sense that it forgets its initial conditions, we also say that the quasi-invariant measure η_∞ is the Yaglom measure. Whenever it exists, we let η_∞^h be the invariant measure of the h-process X_n^h. Under our assumptions, it is now a simple exercise to check that

$$\eta_\infty = \Psi_{M(h)}(\mu) \quad \text{and} \quad \eta_\infty^h := \Psi_h(\eta_\infty) = \Psi_{hM(h)}(\mu).$$

Quantitative convergence estimates of the limiting formulae (2.18) and (2.19) can be derived using the stability properties of the Feynman-Kac models developed in Section 3. For a more thorough discussion on these particle absorption models, we refer the reader to [17, 20, 30, 31, 38, 39].

3

Feynman-Kac Semigroup Analysis

3.1 Introduction

As we mentioned in subsection 1.4, the concentration analysis of particle models is intimately related to the regularity properties of the limiting nonlinear semigroup. Here, we survey some selected topics on the theory of Feynman-Kac semigroups developed in the series of articles [23, 31, 38]. For more recent treatments, we also refer the reader to the books [17, 20].

We begin this section with a discussion on path space models. Subsection 3.2 is concerned with Feynman-Kac historical processes and Backward Markov chain interpretation models. We show that the the n-th marginal measures η_n of a Feynman-Kac model with a reference historical Markov process coincides with the path space measures $\mathbb{Q}_n$ introduced in (1.4).

The second part of subsection 3.2 is dedicated to the proof of the Backward Markov chain formulae (1.12). In subsection 3.3, we analyze the regularity and the semigroup structure of the normalized and unnormalized Feynman-Kac distribution flows η_n and γ_n.

Subsection 3.4 is concerned with the stability properties of the normalized Feynman-Kac distribution flow. In subsection 3.4.1, we present

regularity conditions on the potential functions G_n and on the Markov transitions M_n, under which the Feynman-Kac semigroup forgets exponentially fast its initial condition. Quantitative contraction theorems are provided in subsection 3.4.2.

We illustrate these results with three applications related to time discretization techniques, simulated annealing type schemes, and path space models, respectively.

Subsections 3.5 and 3.6 are concerned with mean field stochastic particle models and local sampling random field models.

3.2 Historical and Backward Models

The historical process associated with some reference Markov chain X_n is defined by the sequence of random paths

$$\mathbf{X}_n = (X_0, \ldots, X_n) \in \mathbf{E_n} := (E_0 \times \ldots \times E_n).$$

Notice that the Markov transitions of the chain $\mathbf{X}_n$ are given for any $\mathbf{x}_{n-1} = (x_0, \ldots, x_{n-1})$ and $\mathbf{y_n} = (y_0, \ldots, y_n) = (\mathbf{y_{n-1}}, y_n)$ by the following formulae

$$\mathbf{M}_n(\mathbf{x}_{n-1}, d\mathbf{y_n}) = \delta_{\mathbf{x}_{n-1}}(d\mathbf{y_{n-1}}) M_n(y_{n-1}, dx_n). \tag{3.1}$$

We consider a sequence of $(0,1]$-valued potential functions $\mathbf{G}_n$ on $\mathbf{E_n}$ whose values only depend on the final state of the paths; that is, we have that

$$\mathbf{G}_n : \mathbf{x}_n = (x_0, \ldots, x_n) \in \mathbf{E_n} \mapsto \mathbf{G}_n(\mathbf{x}_n) = G_n(x_n) \in (0,1] \tag{3.2}$$

with some $(0,1]$-valued potential function G_n on E_n.

We let (γ_n, η_n) represent the Feynman-Kac model associated with the pair $(\mathbf{G}_n, \mathbf{M}_n)$ on the path spaces $\mathbf{E_n}$. By construction, for any function $\mathbf{f}_n$ on $\mathbf{E_n}$, we have

$$\begin{aligned}
\gamma_n(\mathbf{f}_n) &= \mathbb{E}\left(\mathbf{f}_n(\mathbf{X}_n) \prod_{0 \le p < n} \mathbf{G}_p(\mathbf{X}_p) \right) \\
&= \mathbb{E}\left(\mathbf{f}_n(X_0, \ldots, X_n) \prod_{0 \le p < n} G_p(X_p) \right)
\end{aligned}$$

from which we conclude that

$$\gamma_n = \mathcal{Z}_n \mathbb{Q}_n \quad \text{and} \quad \eta_n = \mathbb{Q}_n \tag{3.3}$$

where $\mathbb{Q}_n$ is the Feynman-Kac measure on the path space associated with the pair (G_n, M_n), and defined in (1.4).

We end with the proof of the backward formula (1.12). Using the decomposition

$$\mathbb{Q}_n(d(x_0, \ldots, x_n)) = \frac{\mathcal{Z}_{n-1}}{\mathcal{Z}_n} \mathbb{Q}_{n-1}(d(x_0, \ldots, x_{n-1})) Q_n(x_{n-1}, dx_n)$$

we prove the following formulae

$$\eta_n(dx_n) = \frac{\mathcal{Z}_{n-1}}{\mathcal{Z}_n} \eta_{n-1} Q_n(dx_n) \tag{3.4}$$

$$= \frac{\mathcal{Z}_{n-1}}{\mathcal{Z}_n} \eta_{n-1}(G_{n-1} H_n(., x_n)) \lambda_n(dx_n) \tag{3.5}$$

and

$$\frac{\mathcal{Z}_n}{\mathcal{Z}_{n-1}} = \eta_{n-1} Q_n(\mathbb{1}) = \eta_{n-1}(G_{n-1}).$$

This implies that

$$\frac{d\eta_{n-1}Q_n}{d\eta_n}(x_n) \times \frac{d\eta_{n-2}Q_{n-1}}{d\eta_{n-1}}(x_{n-1}) \times \cdots \times \frac{d\eta_0 Q_1}{d\eta_1}(x_1)$$
$$= \frac{\mathcal{Z}_n}{\mathcal{Z}_{n-1}} \times \frac{\mathcal{Z}_{n-1}}{\mathcal{Z}_{n-2}} \times \cdots \times \frac{\mathcal{Z}_1}{\mathcal{Z}_0} = \mathcal{Z}_n.$$

Using these observations, we readily prove the desired backward decomposition formula.

3.3 Semigroup Models

This subsection is concerned with the semigroup structure and the weak regularity properties of Feynman-Kac models.

Definition 3.1. We denote by

$$\Phi_{p,n}(\eta_p) = \eta_n \quad \text{and} \quad \gamma_p Q_{p,n} = \gamma_n$$

with $0 \leq p \leq n$, the linear and the nonlinear semigroup associated with the unnormalized and the normalized Feynman-Kac measures. For $p = n$, we use the convention $\Phi_{n,n} = Id$, the identity mapping.

Notice that $Q_{p,n}$ has the following functional representation

$$Q_{p,n}(f_n)(x_p) := \mathbb{E}\left(f_n(X_n) \prod_{p \leq q < n} G_q(X_q) \mid X_p = x_p \right).$$

Definition 3.2. We let $G_{p,n}$ and $P_{p,n}$ be the potential functions and the Markov transitions defined by

$$Q_{p,n}(\mathbb{1})(x) = G_{p,n}(x) \quad \text{and} \quad P_{p,n}(f) = \frac{Q_{p,n}(f)}{Q_{p,n}(\mathbb{1})}.$$

We also set

$$g_{p,n} := \sup_{x,y} \frac{G_{p,n}(x)}{G_{p,n}(y)} \quad \text{and} \quad \beta(P_{p,n}) = \sup \operatorname{osc}(P_{p,n}(f)).$$

The r.h.s. supremum is taken as the set of functions $\operatorname{Osc}(E)$. To simplify notation, for $n = p + 1$ we have also set

$$G_{p,p+1} = Q_{p,p+1}(\mathbb{1}) = G_p$$

and sometimes we write g_p instead of $g_{p,p+1}$.

The particle concentration inequalities developed in Section 6 will be expressed in terms of the following parameters.

Definition 3.3. For any $k, l \geq 0$, we also set

$$\tau_{k,l}(n) := \sum_{0 \leq p \leq n} g_{p,n}^k \; \beta(P_{p,n})^l \quad \text{and} \quad \kappa(n) := \sup_{0 \leq p \leq n} (g_{p,n} \beta(P_{p,n})). \tag{3.6}$$

Using the fact that

$$\eta_n(f_n) := \eta_p Q_{p,n}(f_n) / \eta_p Q_{p,n}(1) \tag{3.7}$$

we prove the following formula

$$\Phi_{p,n}(\eta_p) = \Psi_{G_{p,n}}(\eta_p) P_{p,n}$$

for any $0 \leq p \leq n$.

As a direct consequence of (1.22) and (1.23), we quote the following weak regularity property of the Feynman-Kac semigroups.

Proposition 3.1. For $[0,1]$-valued potential function G_n, and any couple of measures ν, μ on the set E s.t. $\mu(G_{p,n}) \wedge \nu(G_{p,n}) > 0$, we have the decomposition

$$\Phi_{p,n}(\mu) - \Phi_{p,n}(\nu) = \frac{1}{\nu(G_{p,n})} \ (\mu - \nu) S_{G_{p,n},\mu} P_{p,n}.$$

In addition, we have the following Lipschitz estimates

$$\|\Phi_{p,n}(\mu) - \Phi_{p,n}(\nu)\|_{\mathrm{tv}} \leq \frac{\|G_{p,n}\|}{\mu(G_{p,n}) \vee \nu(G_{p,n})} \ \beta(P_{p,n}) \ \|\mu - \nu\|_{\mathrm{tv}}$$

and

$$\sup_{\mu,\nu} \|\Phi_{p,n}(\mu) - \Phi_{p,n}(\nu)\|_{\mathrm{tv}} = \beta(P_{p,n}).$$

3.4 Stability Properties

3.4.1 Regularity Conditions

Here, we present one of the simplest quantitative contraction estimates known for the normalized Feynman-Kac semigroups $\Phi_{p,n}$. We consider the following regularity conditions.

$\mathbf{H_m(G,M)}$ There exists some integer $m \geq 1$, such that for any $n \geq 0$, and any $((x,x'),A) \in (E_n^2 \times \mathcal{E}_n)$ and any $n \geq 0$ we have

$$M_{n,n+m}(x,A) \leq \chi_m \ M_{n,n+m}(x',A) \quad \text{and} \quad g = \sup_{n \geq 0} g_n < \infty$$

for some finite parameters $\chi_m, g < \infty$, and some integer $m \geq 1$.

$\mathbf{H_0(G,M)}$

$$\rho := \sup_{n \geq 0} (g_n \beta(M_{n+1})) < 1 \quad \text{and} \quad g = \sup_n g_n < \infty. \tag{3.8}$$

Both conditions are related to the stability properties of the reference Markov chain model X_n with probability transition M_n. This

implies that the chain X_n tends to merges exponentially fast the random states starting from *any* two different locations.

One natural strategy to obtain some useful quantitative contraction estimates for the Markov transitions $P_{p,n}$ is to write this transition in terms of the composition of Markov transitions.

Lemma 3.2. For any $0 \leq p \leq q \leq n$, we have

$$P_{p,n} = R^{(n)}_{p,q} P_{q,n} \quad \text{and} \quad P_{p,n} = R^{(n)}_{p+1} R^{(n)}_{p+2} \ldots R^{(n)}_{n-1} R^{(n)}_{n}$$

with the triangular array of Markov transitions $R^{(n)}_{p,q}$ and $(R^{(n)}_{q})_{1 \leq q \leq n}$ defined by

$$R^{(n)}_{p,q}(f) := \frac{Q_{p,q}(G_{q,n} f)}{Q_{p,q}(G_{q,n})} = \frac{P_{p,q}(G_{q,n} f)}{P_{p,q}(G_{q,n})}$$

and

$$R^{(n)}_{p}(f) = \frac{Q_{p}(G_{p,n} f)}{Q_{p}(G_{p,n})} = \frac{M_{p}(G_{p,n} f)}{M_{p}(G_{p,n})}.$$

In addition, for any $0 \leq p \leq q \leq n$ we have

$$\beta(R^{(n)}_{p,q}) \leq g_{q,n} \beta(P_{p,q}) \quad \text{and} \quad \log g_{p,n} \leq \sum_{p \leq q < n} (g_q - 1) \beta(P_{p,q}). \tag{3.9}$$

Proof:
Using the decomposition

$$Q_{p,n}(f) = Q_{p,q}(Q_{q,n}(f)) = Q_{p,q}(Q_{q,n}(1)\ P_{q,n}(f))$$

we can easily check the first assertion. The left-hand side (l.h.s.) inequality in (3.9) is a direct consequence of (1.26). Using (1.9), the proof of the r.h.s. inequality in (3.9) is based on the fact that

$$\begin{aligned} \frac{G_{p,n}(x)}{G_{p,n}(y)} &= \frac{\prod_{p \leq q < n} \Phi_{p,q}(\delta_x)(G_q)}{\prod_{p \leq q < n} \Phi_{p,q}(\delta_y)(G_q)} \\ &= \exp\left\{ \sum_{p \leq q < n} (\log \Phi_{p,q}(\delta_x)(G_q) - \log \Phi_{p,q}(\delta_y)(G_q)) \right\}. \end{aligned}$$

Using the fact that

$$\log y - \log x = \int_0^1 \frac{(y-x)}{x + t(y-x)} dt$$

for any positive numbers x, y, we prove that

$$\frac{G_{p,n}(x)}{G_{p,n}(y)}$$

$$= \exp\left\{ \sum_{p \leq q < n} \int_0^1 \frac{(\Phi_{p,q}(\delta_x)(G_q) - \Phi_{p,q}(\delta_y)(G_q))}{\Phi_{p,q}(\delta_y)(G_q) + t(\Phi_{p,q}(\delta_x)(G_q) - \Phi_{p,q}(\delta_y)(G_q))} dt \right\}$$

$$\leq \exp\left\{ \sum_{p \leq q < n} \widetilde{g}_q \times (\Phi_{p,q}(\delta_x)(\widetilde{G}_q) - \Phi_{p,q}(\delta_y)(\widetilde{G}_q)) \right\}$$

with

$$\widetilde{G}_q := G_q / \mathrm{osc}(G_q) \quad \text{and} \quad \widetilde{g}_q := \mathrm{osc}(G_q)/\inf G_q \leq g_q - 1.$$

We end the proof of the desired estimates using (1.26), and Proposition 3.1. This completes the proof of the lemma. ∎

3.4.2 Quantitative Contraction Theorems

This subsection is mainly concerned with the proof of two contraction theorems, which can be derived under the couple of regularity conditions presented in subsection 3.4.1.

Theorem 3.3. We assume that condition $\mathbf{H_m}(\mathbf{G}, \mathbf{M})$ is satisfied for some finite parameters $\chi_m, g < \infty$, and some integer $m \geq 1$. In this situation, we have the uniform estimates

$$\sup_{0 \leq p \leq n} g_{p,n} \leq \chi_m g^m \quad \text{and} \quad \sup_{p \geq 0} \beta(P_{p,p+km}) \leq (1 - g^{-(m-1)} \chi_m^{-2})^k. \tag{3.10}$$

In addition, for any couple of measures $\nu, \mu \in \mathcal{P}(E_p)$, and for any $f \in \mathrm{Osc}(E_n)$ we have the decomposition

$$|[\Phi_{p,n}(\mu) - \Phi_{p,n}(\nu)](f)| \leq \rho_m (1 - \kappa_m)^{(n-p)/m} |(\mu - \nu) D_{p,n,\mu}(f)| \tag{3.11}$$

for some function $D_{p,n,\mu}(f) \in \mathrm{Osc}(E_p)$ whose values only depends on the parameters (p,n,μ), and some parameters $\rho_m < \infty$ and $\kappa_m \in]0,1]$ such that

$$\rho_m \leq \chi_m g^m (1 - g^{-(m-1)}\chi_m^{-2})^{-1} \quad \text{and} \quad \kappa_m \geq g^{-(m-1)}\chi_m^{-2}. \tag{3.12}$$

Proof:

For any nonnegative function f, we notice that

$$\begin{aligned} R^{(n)}_{p,p+m}(f)(x) &= \frac{Q_{p,p+m}(G_{p+m,n}f)(x)}{Q_{p,p+m}(G_{p+m,n})(x)} \\ &\geq g^{-(m-1)}\chi_m^{-2}\frac{M_{p,p+m}(G_{p+m,n}f)(x')}{M_{p,p+m}(G_{p+m,n})(x')} \end{aligned}$$

and for any $p + m \leq n$

$$\frac{G_{p,n}(x)}{G_{p,n}(x')} = \frac{Q_{p,p+m}(G_{p+m,n})(x)}{Q_{p,p+m}(G_{p+m,n})(x')} \leq g^m \frac{M_{p,p+m}(G_{p+m,n})(x)}{M_{p,p+m}(G_{p+m,n})(x')} \leq \chi_m g^m.$$

For $p \leq n \leq p + m$, this upper bound remains valid. We conclude that

$$G_{p,n}(x) \leq \chi_m g^m \; G_{p,n}(x') \quad \text{and} \quad \beta(R^{(n)}_{p,p+m}) \leq 1 - g^{-(m-1)}\chi_m^{-2}.$$

In the same way as above, we have

$$n = km \Rightarrow P_{p,p+km} = R^{(n)}_{p,p+m} R^{(n)}_{p+m,p+2m} \cdots R^{(n)}_{p+(k-1)m,p+km}$$

and

$$\beta(P_{p,p+km}) \leq \prod_{1 \leq l \leq k} \beta(R^{(n)}_{p+(l-1)m,p+lm}) \leq (1 - g^{-(m-1)}\chi_m^{-2})^k.$$

This ends the proof of (3.10).

The proof of (3.11) is based on the decomposition

$$(\mu - \nu)S_{G_{p,n},\mu}P_{p,n}(f) = \beta(S_{G_{p,n},\mu}P_{p,n}) \times (\mu - \nu)D_{p,n,\mu}(f)$$

with

$$D_{p,n,\mu}(f) := S_{G_{p,n},\mu}P_{p,n}(f)/\beta(S_{G_{p,n},\mu}P_{p,n}).$$

On the other hand, we have

$$S_{G_{p,n},\mu}(x,y) \geq (1 - \|G_{p,n}\|) \Rightarrow \beta(S_{G_{p,n},\mu}) \leq \|G_{p,n}\|$$

and

$$\beta(S_{G_{p,n},\mu})/\nu(G_{p,n}) \leq g_{p,n} \leq \chi_m g^m.$$

Finally, we observe that

$$\beta(P_{p,n}) \leq \beta(P_{p,p+\lfloor (n-p)/m \rfloor})$$

from which we conclude that

$$\frac{1}{\nu(G_{p,n})}\ \beta(S_{G_{p,n},\mu}P_{p,n}) \leq \chi_m g^m \beta(P_{p,p+\lfloor (n-p)/m \rfloor}).$$

The end of the proof is now a direct consequence of the contraction estimate (3.10). This ends the proof of the theorem. ∎

Theorem 3.4. We assume that condition $\mathbf{H_0}(\mathbf{G},\mathbf{M})$ is satisfied for some $\rho < 1$. In this situation, for any couple of measures $\nu, \mu \in \mathcal{P}(E_p)$, and for any $f \in \mathrm{Osc}(E_n)$ we have the decomposition

$$|[\Phi_{p,n}(\mu) - \Phi_{p,n}(\nu)](f)| \leq \rho^{(n-p)} |(\mu - \nu) D_{p,n,\mu}(f)|$$

for some function $D_{p,n,\mu}(f) \in \mathrm{Osc}(E_p)$, whose values only depends on the parameters (p,n,μ). In addition, for any $0 \leq p \leq n$, we have the estimates

$$\beta(P_{p,n}) \leq \rho^{n-p} \quad \text{and} \quad g_{p,n} \leq \exp\left((g-1)\ (1-\rho^{n-p})/(1-\rho)\right).$$

Proof. Using Proposition 3.1, and recalling that $\beta(S_{G_{n-1},\mu}) \leq \|G_{n-1}\|$, we readily prove that

$$\begin{aligned} |[\Phi_n(\mu) - \Phi_n(\nu)](f)| &\leq g_{n-1}\beta(M_n)|(\mu-\nu)D_{n,\mu}(f)| \\ &\leq \rho|(\mu-\nu)D_{n,\mu}(f)| \end{aligned}$$

with the function

$$D_{n,\mu}(f) = S_{G_{n-1},\mu}M_n(f)/\beta(S_{G_{n-1},\mu}M_n) \in \mathrm{Osc}(E_{n-1}).$$

Now, we can prove the theorem by induction on the parameter $n \geq p$. For $n = p$, the desired result follows from the above discussion. Suppose we have

$$|[\Phi_{p,n-1}(\mu) - \Phi_{p,n-1}(\nu)](f)| \leq \rho^{(n-p-1)} |(\mu - \nu) D_{p,n-1,\mu}(f)|$$

for any $f \in \mathrm{Osc}(E_{n-1})$, and some functions $D_{p,n-1,\mu}(f) \in \mathrm{Osc}(E_p)$. In this case, we have

$$\begin{aligned}|[\Phi_n(\Phi_{p,n-1}(\mu)) - \Phi_n(\Phi_{p,n-1}(\nu))](f)| \\ \leq g_{n-1}\beta(M_n)|(\Phi_{p,n-1}(\mu) - \Phi_{p,n-1}(\nu))D_{n,\Phi_{p,n-1}(\mu)}(f)|\end{aligned}$$

for any $f \in \mathrm{Osc}(E_n)$, with $D_{n,\Phi_{p,n-1}(\mu)}(f) \in \mathrm{Osc}(E_{n-1})$.

Under our assumptions, we conclude that

$$|[\Phi_n(\Phi_{p,n-1}(\mu)) - \Phi_n(\Phi_{p,n-1}(\nu))](f)| \leq \rho^{(n-p)}|(\mu - \nu)D_{p,n,\mu}(f)|$$

with the function

$$D_{p,n,\mu}(f) := D_{p,n-1,\mu}(D_{n,\Phi_{p,n-1}(\mu)}(f)) \in \mathrm{Osc}(E_p).$$

The proof of the second assertion is a direct consequence of Proposition 3.1, and Lemma 3.2. This ends the proof of the theorem. ∎

Corollary 3.5. Under any of the conditions $\mathbf{H_m}(\mathbf{G},\mathbf{M})$, with $m \geq 0$, the functions $\tau_{k,l}$ and κ defined in (3.6) are uniformly bounded; that is, for any $k,l \geq 1$ we have that

$$\overline{\tau}_{k,l}(m) := \sup_{n\geq 0}\sup_{0\leq p\leq n} \tau_{k,l}(n) < \infty \quad \text{and} \quad \overline{\kappa}(m) := \sup_{n\geq 0}\kappa(n) < \infty.$$

In addition, for any $m \geq 1$, we have

$$\begin{aligned}\overline{\kappa}(m) &\in [1, \chi_m g^m] \\ \overline{\tau}_{k,l}(m) &\leq m(\chi_m g^m)^k \ /(1 - (1 - g^{-(m-1)}\chi_m^{-2})^l)\end{aligned}$$

and for $m = 0$, we have the estimates

$$\begin{aligned}\overline{\kappa}(0) &\leq \exp((g-1)/(1-\rho)) \\ \overline{\tau}_{k,l}(0) &\leq \exp(k(g-1)/(1-\rho))/(1-\rho^l).\end{aligned}$$

3.4.3 Some Illustrations

We illustrate the regularity conditions presented in subsection 3.4.1 with three different types of Feynman-Kac models, related to time

discretization techniques, simulated annealing type schemes, and path space models, respectively.

Of course, a complete analysis of the regularity properties of the 20 Feynman-Kac application models presented in section 2 would lead to too long of a discussion.

In some instances, the regularity conditions stated in subsection 3.4.1 can be directly translated into regularity properties of the reference Markov chain model and the adaptation potential function.

In other instances, the regularity properties of the Feynman-Kac semigroup depend on some important tuning parameters, including discretization time steps, and cooling schedules in simulated annealing time models. In subsection 3.4.3.1, we illustrate the regularity property $\mathbf{H_0}(\mathbf{G},\mathbf{M})$ stated in (3.8) in the context of a time discretization model with geometric style clocks introduced in (2.17). In subsection 3.4.3.2, we present some tools to tune the cooling parameters of the annealing model discussed in (2.2), so that the resulting semigroups are exponentially stable.

For degenerate indicator style functions, we can use a one-step integration technique to transform the model into a Feynman-Kac model on smaller state spaces with positive potential functions. In terms of particle absorption models, this technique allows us to turn a hard obstacle model into a soft obstacle particle model. Further details on this integration technique can be found in [17, 19].

Last, but not least, in some important applications, including Feynman-Kac models on path spaces, the limiting semigroups are unstable, in the sense that they do not forget their initial conditions. Nevertheless, in some situations it is still possible to control uniformly in time the quantities $g_{p,n}$ discussed in subsection 3.3.

3.4.3.1 Time Discretization Models

We consider that the potential functions G_n and Markov transitions M_n are given by (2.17), for some non negative function V_n, some positive parameter λ_n and some Markov transition K_n s.t.

$$\beta(K_n) \leq \kappa_n < 1 \quad h \leq h_n = (1-\kappa_n)/[v_{n-1}+\alpha] \quad \text{and} \quad \lambda_n \in]0, 1/h]$$

with $v_{n-1} := \mathrm{osc}(V_{n-1})$, and for some $\alpha > 0$. We also assume that $v = \sup_n v_n < \infty$.

In this situation, for any $\lambda_n \in [\frac{1}{h_n}, \frac{1}{h}]$, we have

$$\begin{aligned} g_{n-1}\beta(M_n) &\leq e^{v_{n-1}h}(1 - \lambda_n h\ (1 - \kappa_n)) \\ &\leq e^{-h(\lambda_n(1-\kappa_n)-v_{n-1})} \leq e^{-\alpha h} \end{aligned}$$

from which we conclude that $\mathbf{H_0}(\mathbf{G}, \mathbf{M})$ is met with

$$g = \sup_n g_n \leq e^{hv} \quad \text{and} \quad \rho \leq e^{-\alpha h}.$$

3.4.3.2 Interacting Simulated Annealing Model

We consider the Feynman-Kac annealing model discussed in (2.2). We further assume that

$$K_n^{k_n}(x, dy) \geq \epsilon_n\ \nu_n(y)$$

for some $k_n \geq 1$, some $\epsilon_n > 0$, and some measure ν_n.

In this situation, we have

$$\mathcal{M}_{n,\beta_n}^{k_n}(x, dy) \geq K_n^{k_n}(x, dy)\ e^{-\beta_n k_n v} \geq \epsilon_n\ e^{-\beta_n k_n v}\ \nu_n(dy)$$

with $v := \mathrm{osc}(V)$. If we choose $m_n = k_n l_n$, this implies that

$$\beta(M_n) = \beta(\mathcal{M}_{n,\beta_n}^{m_n}) \leq \beta(\mathcal{M}_{n,\beta_n}^{k_n})^{l_n} \leq (1 - \epsilon_n\ e^{-\beta_n k_n v})^{l_n}.$$

Therefore, for any given $\rho' \in]0, 1[$ we can chose l_n such that

$$l_n \geq \frac{\log(1/\rho') + v(\beta_n - \beta_{n-1})}{\log 1/(1 - \epsilon_n\ e^{-\beta_n k_n v})}$$

so that

$$g_{n-1}\beta(M_n) \leq e^{v(\beta_n - \beta_{n-1})} \times (1 - \epsilon_n\ e^{-\beta_n k_n v})^{l_n} \leq \rho' \Rightarrow \rho \leq \rho'.$$

For any function $\beta : x \in [0, \infty[\mapsto \beta(x)$, with a decreasing derivative $\beta'(x)$ s.t. $\lim_{x\to\infty} \beta'(x) = 0$ and $\beta'(0) < \infty$, we also notice that

$$g = \sup_{n\geq 0} g_n \leq \sup_{n\geq 0} e^{v\beta'_n} \leq e^{v\beta'(0)}.$$

3.4.3.3 Historical Processes

We return to the historical Feynman-Kac models introduced in subsection 3.2. Using the equivalence principle (3.3), we have proved that the n-time marginal models associated with a Feynman-Kac model on path space coincide with the original Feynman-Kac measure (1.4).

We write $\mathbf{Q}_{p,n}$ and $\mathbf{P}_{p,n}$ the Feynman-Kac semigroups defined as $Q_{p,n}$ and $P_{p,n}$, by replacing (G_n, M_n) by $(\mathbf{G}_n, \mathbf{M}_n)$. By construction, we have

$$\mathbf{Q}_{p,n}(\mathbb{1})(\mathbf{x}_p) = Q_{p,n}(\mathbb{1})(x_p)$$

for any $\mathbf{x}_p = (x_0, \ldots, x_p) \in E_p$. Therefore, if we set

$$\mathbf{G}_{p,n}(\mathbf{x}_p) := \mathbf{Q}_{p,n}(\mathbb{1})(\mathbf{x}_p)$$

then we find that

$$\mathbf{g}_{p,n} := \sup_{\mathbf{x}_p, \mathbf{y}_p} \frac{\mathbf{G}_{p,n}(\mathbf{x}_p)}{\mathbf{G}_{p,n}(\mathbf{y}_p)} = \sup_{x_p, y_p} \frac{G_{p,n}(x_p)}{G_{p,n}(y_p)} = g_{p,n}.$$

On the other hand, we cannot expect the Dobrushin's ergodic coefficient of the historical process semigroup to decrease, but we always have $\beta(\mathbf{P}_{p,n}) \leq 1$.

In summary, when the reference Markov chain X_n satisfies the condition $\mathbf{H_m}(\mathbf{G}, \mathbf{M})$ stated in the beginning of subsection 3.4.1, for some $m \geq 1$, we always have the estimates

$$\mathbf{g}_{p,n} \leq \chi_m g^m \quad \text{and} \quad \beta(\mathbf{P}_{p,n}) \leq 1 \tag{3.13}$$

and

$$\tau_{k,l}(n) \leq (n+1)(\chi_m g^m)^k \quad \text{and} \quad \kappa(n) \leq \chi_m g^m \tag{3.14}$$

with the functions $\tau_{k,l}$, and κ introduced in (3.6)

We end with some Markov chain Monte Carlo techniques often used in practice to stabilize the genealogical tree based approximation model. To describe with some precision this stochastic method, we consider the Feynman-Kac measures $\eta_n \in \mathcal{P}(\mathbf{E_n})$ associated with the potential function $\mathbf{G}_n$ and the Markov transitions $\mathbf{M}_n$ of the historical process defined respectively in (3.2) and in (3.1). We notice that the measures η_n satisfy the updating-prediction equation

$$\eta_n = \Psi_{\mathbf{G}_n}(\eta_n)\mathbf{M}_n.$$

This equation on the set of measures on path spaces is unstable, in the sense that its initial condition is always kept in mind by the historical Markov transitions $\mathbf{M_n}$. One idea to stabilize this system is to incorporate an additional Markov chain Monte Carlo move at every time step. More formally, let us suppose that we have a dedicated Markov chain Monte Carlo transition K_n from the set $\mathbf{E_n}$ into itself, and such that

$$\eta_n = \eta_n K_n.$$

In this situation, we also have that

$$\eta_n = \Psi_{\mathbf{G}_n}(\eta_n)\mathbf{M}'_n \quad \text{with } \mathbf{M}'_{\mathbf{n}} := \mathbf{M}_n K_n. \tag{3.15}$$

By construction, the mean field particle approximation of the equation (3.15) is a genealogical tree type evolution model with path space particles on the state spaces $\mathbf{E_n}$. The updating-selection transitions are related to the potential function $\mathbf{G_n}$ on the state spaces $\mathbf{E_n}$, and the mutation-exploration mechanisms from $\mathbf{E_n}$ into $\mathbf{E_{n+1}}$ are dictated by the Markov transitions $\mathbf{M'_{n+1}}$.

Notice that this mutation transition is decomposed into two different stages. First, we extend the selected path-valued particles with an elementary move according to the Markov transition M_n. Then, from each of these extended paths, we perform a Markov chain Monte Carlo sample according to the Markov transition K_n.

3.5 Mean Field Particle Models

With the exception of some very special cases, the measures η_n cannot be represented in a closed form, even on infinite dimensional state-spaces. Their numerical estimation using deterministic type grid approximations requires extensive calculations, and they rarely cope with high-dimensional problems. In the same vein, harmonic type approximation schemes, or related linearization style techniques such as the extended Kalman filter often provide poor estimation results for highly nonlinear models. In contrast with these conventional techniques, mean field particle models can be thought of as a stochastic adaptive grid approximation scheme. These advanced Monte Carlo methods take advantage of the nonlinearities of the model, to design an interacting selection-recycling mechanism.

Formally speaking, discrete generation mean field particle models are based on the fact that the flow of probability measures η_n satisfies a non linear evolution equation of the following form

$$\eta_{n+1}(dy) = \int \eta_n(dx) K_{n+1,\eta_n}(x, dy) \tag{3.16}$$

for some collection of Markov transitions $K_{n+1,\eta}$, indexed by the time parameter $n \geq 0$ and the set of probability measures $\mathcal{P}(E_n)$.

The choice of the McKean transitions K_{n+1,η_n} is not unique. For instance, we can choose

$$K_{n+1,\eta_n}(x, dy) = \Phi_{n+1}(\eta_n)(dy)$$

and more generally

$$\begin{aligned} &K_{n+1,\eta_n}(x, dy) \\ &\quad = \epsilon(\eta_n) G_n(x)\ M_{n+1}(x, dy) + (1 - \epsilon(\eta_n) G_n(x))\ \Phi_{n+1}(\eta_n)(dy) \end{aligned}$$

for any $\epsilon(\eta_n)$ s.t. $\epsilon(\eta_n) G_n(x) \in [0, 1]$. Note that we can define sequentially a Markov chain sequence $(\overline{X}_n)_{n \geq 0}$ such that

$$\mathbb{P}(\overline{X}_{n+1} \in dx \mid \overline{X}_n) = K_{n+1,\eta_n}(\overline{X}_n, dx) \quad \text{with Law}(\overline{X}_n) = \eta_n.$$

From practical point of view, this Markov chain can be seen as a perfect sampler of the flow of the distributions (3.16) of the random states $\overline{X}_n$. For a more thorough discussion on these nonlinear Markov chain models, we refer the reader to section 2.5 in the book [17].

The mean field particle interpretation of this nonlinear measure valued model is the E_n^N-valued Markov chain

$$\xi_n = (\xi_n^1, \xi_n^2, \ldots, \xi_n^N) \in E_n^N$$

with elementary transitions defined as

$$\mathbb{P}(\xi_{n+1} \in dx \mid \xi_n) = \prod_{i=1}^{N} K_{n+1,\eta_n^N}(\xi_n^i, dx^i) \tag{3.17}$$

with

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

In the above displayed formula, dx stands for an infinitesimal neighborhood of the point $x = (x^1, \ldots, x^N) \in E_{n+1}^N$. The initial system ξ_0 consists of N independent and identically distributed random variables with common law η_0.

We let $\mathcal{G}_n^N := \sigma(\xi_0, \ldots, \xi_n)$ be the natural filtration associated with the N-particle approximation model defined above.

The particle model associated with the parameter $\epsilon(\eta_n) = 1$ coincides with the genetic type stochastic algorithm presented in subsection 1.4.

Furthermore, using the equivalence principles (3.3) presented in subsection 3.2, we can check that the genealogical tree model discussed above coincides with the mean field N-particle interpretation of the Feynman-Kac measures (γ_n, η_n) associated with the pair $(\mathbf{G}_n, \mathbf{M}_n)$ on the path spaces $\mathbf{E_n}$. In this context, we recall that $\eta_n = \mathbb{Q}_n$, and the N-particle approximation measures are given by

$$\eta_n^N := \frac{1}{N} \sum_{i=1}^{N} \delta_{(\xi_{0,n}^i, \xi_{1,n}^i, \ldots, \xi_{n,n}^i)} \in \mathcal{P}(\mathbf{E_n}) = \mathcal{P}(E_0 \times \ldots \times E_n). \tag{3.18}$$

3.6 Local Sampling Errors

The local sampling errors induced by the mean field particle model (3.17) are expressed in terms of the empirical random field sequence V_n^N defined by

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

Notice that V_{n+1}^N is alternatively defined by the following stochastic perturbation formulae

$$\eta_{n+1}^N = \Phi_{n+1}(\eta_n^N) + \frac{1}{\sqrt{N}} V_{n+1}^N. \tag{3.19}$$

For $n = 0$, we also set

$$V_0^N = \sqrt{N}[\eta_0^N - \eta_0] \Leftrightarrow \eta_0^N = \eta_0 + \frac{1}{\sqrt{N}} V_0^N.$$

In this interpretation, the N-particle model can also be interpreted as a stochastic perturbation of the limiting system

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

It is rather elementary to check that

$$\begin{aligned}\mathbb{E}(V_{n+1}^N(f)|\ \mathcal{G}_n^N.) &= 0\\ \mathbb{E}(V_{n+1}^N(f)^2|\ \mathcal{G}_n^N) &= \eta_n^N[K_{n+1,\eta_n^N}(f - K_{n+1,\eta_n^N}(f))^2].\end{aligned}$$

Definition 3.4. We denote by σ_n^2 the uniform local variance parameter given by

$$\sigma_n^2 := \sup \mu(K_{n,\mu}[f_n - K_{n,\mu}(f_n)]^2) \leq 1. \tag{3.20}$$

In the above displayed formula the supremum is taken over all functions $f_n \in \mathrm{Osc}(E_n)$, and all probability measures μ on E_n, with $n \geq 1$. For $n = 0$, we set

$$\sigma_0^2 = \sup_{f_0 \in \mathrm{Osc}(E_0)} \eta_0([f_0 - \eta_0(f_0)]^2) \leq 1.$$

We close with a brief discussion on these uniform local variance parameters in the context of continuous time discretization models. When the discrete time model $K_{n,\mu} = K_{n,\mu}^{(h)}$ comes from a discretization of the continuous time model with time step $\Delta t = h(\leq 1)$, we often have that

$$K_{n,\mu} = Id + hL_{n,\mu} + \mathrm{O}(h^2) \tag{3.21}$$

for some infinitesimal generator $L_{n,\mu}$. In this situation, we also have that

$$L_{K_{n,\mu}} := K_{n,\mu} - Id = hL_{n,\mu} + \mathrm{O}(h^2).$$

For any Markov transition K, we notice that

$$\begin{aligned}K([f - K(f)]^2) &= K(f^2) - K(f)^2\\ &= L_K(f^2) - 2fL_K(f) - (L_K(f))^2\\ &= \Gamma_{L_K}(f,f) - (L_K(f))^2\end{aligned}$$

with the "carré du champ" function $\Gamma_{L_K}(f,f)$ defined for any $x \in E$ by

$$\begin{aligned}\Gamma_{L_K}(f,f)(x) &= L_K([f - L_K(f)(x)]^2)(x) \\ &= L_K(f^2)(x) - 2f(x)\ L_K(f)(x).\end{aligned}$$

When $K = K_{n,\mu}$ and $L_{K_{n,\mu}} = hL_{n,\mu} + \mathrm{O}(h^2)$, we find that

$$\begin{aligned}\mu[K_{n,\mu}[f_n - K_{n,\mu}(f_n)]^2] &= \mu\Gamma_{L_{K_{n,\mu}}}(f,f)h - h^2\mu(L_{K_{n,\mu}}(f)^2) \\ &= h\mu(\Gamma_{L_{n,\mu}}(f,f)) + \mathrm{O}(h^2).\end{aligned}$$

4

Empirical Processes

4.1 Introduction

The aim of this section is to review some more or less well known stochastic techniques for analyzing the concentration properties of empirical processes associated with independent random sequences. Our discussion begins by providing some basic definitions on empirical processes associated with sequences of independent random variables on general measurable state spaces. In subsection 4.2, we state and comment on the main results of this section. Subsection 4.2.1 is concerned with finite marginal models. In subsection 4.2.2, we extend these results at the level of the empirical processes. In addition to the fact that the concentration inequalities for empirical processes hold for the supremum of empirical processes over infinite collection of functions, these inequalities are more crude with greater constants than the ones for marginal models. These two subsections also contain two new perturbation theorems that apply to nonlinear functional of empirical processes. The proofs of these theorems combine Orlicz's norm techniques, Kintchine's type inequalities, maximal inequalities, as well as Laplace-Cramèr-Chernov estimation methods. These four complementary methodologies are presented in subsections 4.3–4.6, respectively.

Let $(\mu^i)_{i\geq 1}$ be a sequence of probability measures on a given measurable state space $(E,\mathcal{E})$. During the further development of this subsection, we fix an integer $N \geq 1$. To clarify the presentation, we slightly abuse the notation and denote by

$$m(X) = \frac{1}{N}\sum_{i=1}^{N} \delta_{X^i} \quad \text{and} \quad \mu = \frac{1}{N}\sum_{i=1}^{N} \mu^i$$

the N-empirical measure associated with a collection of independent random variables $X = (X^i)_{i\geq 1}$, with respective distributions $(\mu^i)_{i\geq 1}$, and the N-averaged measure associated with the sequence of measures $(\mu^i)_{i\geq 1}$, respectively. We also consider the empirical random field sequences

$$V(X) = \sqrt{N}\ (m(X) - \mu).$$

We also set

$$\sigma(f)^2 := \mathbb{E}(V(X)(f)^2) = \frac{1}{N}\sum_{i=1}^{N} \mu^i([f - \mu^i(f)]^2). \tag{4.1}$$

Remark 4.1. The rather abstract models presented above can be used to analyze the local sampling random fields models associated with a mean field particle model discussed in subsection 3.6.

To be more precise, given the information on the N-particle model at time $(n-1)$, the sequences of random variables ξ_n^i are independent random sequences with a distribution that depends on the current state ξ_{n-1}^i. That is, at any given fixed time horizon n and given $\mathcal{G}_{n-1}^N$, we have

$$X^i = \xi_n^i \in E = E_n \quad \text{and} \quad \mu^i(dx) := K_{n,\eta_{n-1}^N}(\xi_{n-1}^i, dx). \tag{4.2}$$

In this case, we find that

$$m(X) = \eta_n^N \quad \text{and} \quad V(X) = V_n^N$$

and

$$\begin{aligned}\sigma(f)^2 &= \mathbb{E}(V_{n+1}^N(f)^2 \mid \mathcal{G}_n^N) \\ &= \eta_n^N[K_{n+1,\eta_n^N}(f - K_{n+1,\eta_n^N}(f))^2].\end{aligned}$$

Let $\mathcal{F}$ be a given collection of measurable functions $f : E \to \mathbb{R}$ such that $\|f\| \leq 1$. We associate with $\mathcal{F}$ the Zolotarev seminorm on $\mathcal{P}(E)$ defined by

$$\|\mu - \nu\|_{\mathcal{F}} = \sup\{|\mu(f) - \nu(f)|; f \in \mathcal{F}\},$$

(see, for instance, [83]). No generality is lost and much convenience is gained by supposing that the unit and the null functions are $f = \mathbb{1}$ and $f = 0 \in \mathcal{F}$. Furthermore, to avoid some unnecessary technical measurability questions, we shall also suppose that $\mathcal{F}$ is separable, in the sense of Doob.

We measure the size of a given class $\mathcal{F}$ in terms of the covering numbers $N(\varepsilon, \mathcal{F}, \mathbb{L}_2(\mu))$ defined as the minimal number of $\mathbb{L}_2(\mu)$-balls of radius $\varepsilon > 0$ needed to cover $\mathcal{F}$. We shall also use the following uniform covering numbers and entropies.

We end with the last of the notations to be used in this section dedicated to empirical processes concentration inequalities.

Definition 4.1. By $\mathcal{N}(\varepsilon, \mathcal{F})$, $\varepsilon > 0$, and by $I(\mathcal{F})$ we denote the uniform covering numbers and entropy integral given by

$$\mathcal{N}(\varepsilon, \mathcal{F}) = \sup\{\mathcal{N}(\varepsilon, \mathcal{F}, \mathbb{L}_2(\eta)); \eta \in \mathcal{P}(E)\}$$
$$I(\mathcal{F}) = \int_0^2 \sqrt{\log\left(1 + \mathcal{N}(\varepsilon, \mathcal{F})\right)}\, d\varepsilon.$$

The concentration inequalities stated in this subsection are expressed in terms of the inverse of the functions defined below.

Definition 4.2. We let (ϵ_0, ϵ_1) be the functions on $\mathbb{R}_+$ defined by

$$\epsilon_0(\lambda) = \frac{1}{2}(\lambda - \log(1 + \lambda)) \quad \text{and} \quad \epsilon_1(\lambda) = (1 + \lambda)\log(1 + \lambda) - \lambda.$$

Rather crude estimates can be derived using the following upper bounds

$$\epsilon_0^{-1}(x) \leq 2(x + \sqrt{x}) \quad \text{and} \quad \epsilon_1^{-1}(x) \leq \frac{x}{3} + \sqrt{2x}.$$

A proof of these elementary inequalities and refined estimates can be found in the recent article [40].

4.2 Statement of the Main Results

4.2.1 Finite Marginal Models

The main result of this subsection is a quantitative concentration inequality for the finite marginal models

$$f \mapsto V(X)(f) = \sqrt{N}(m(X) - \mu(f)).$$

In the following theorem, we provide Kintchine's type mean error bounds, and related Orlicz norm estimates. The detailed proofs of these results are postponed to subsection 4.4. The last quantitative concentration inequality is a direct consequence of (4.5), and it is proved in Remark 4.6.

Theorem 4.1. For any integer $m \geq 1$, and any measurable function f we have the $\mathbb{L}_m$-mean error estimates

$$\mathbb{E}(|V(X)(f)|^m)^{1/m} \leq b(m)(\mathrm{osc}(f) \wedge [2\ \mu(|f|^{m'})^{1/m'}]) \tag{4.3}$$

and

$$\mathbb{E}(|V(X)(f)|^m)^{\frac{1}{m}} \leq\ 6b(m)^2 \max\left(\sqrt{2}\sigma(f), \left[\frac{2\sigma(f)^2}{N^{\frac{m'}{2}-1}}\right]^{1/m'}\right)$$

with the smallest even integer $m' \geq m$, and the collection of constants $b(m)$ defined in (1.27).

In particular, for any $f \in \mathrm{Osc(E)}$, we have

$$\pi_\psi(V(X)(f)) \leq \sqrt{3/8} \tag{4.4}$$

and for any N s.t. $2\sigma^2(f)N \geq 1$ we have

$$\mathbb{E}(|V(X)(f)|^m)^{\frac{1}{m}} \leq 6\sqrt{2}b(m)^2\sigma(f). \tag{4.5}$$

In addition, the probability of the event

$$|V(X)(f)| \leq 6\sqrt{2}\sigma(f)(1 + \epsilon_0^{-1}(x))$$

is greater than $1 - e^{-x}$, for any $x \geq 0$.

In subsection 4.3.2, dedicated to concentration properties of random variables Y with finite Orlicz norms $\pi_\psi(Y) < \infty$, we shall prove that the probability of the event

$$Y \leq \pi_\psi(Y) \ \sqrt{y + \log 2}$$

is greater than $1 - e^{-y}$, for any $y \geq 0$ (cf. lemma 4.6). This implies that the probability of the events

$$|V(X)(f)| \leq \frac{1}{2} \ \sqrt{3(x + \log 2)/2}$$

is greater than $1 - e^{-x}$, for any $x \geq 0$.

Our next objective is to derive concentration inequalities for the nonlinear functional of the empirical random field $V(X)$. To introduce these objects precisely, we need another around of notation.

For any measure ν, and any sequence of measurable functions $f = (f_1, \ldots, f_d)$, we write

$$\nu(f) := [\nu(f_1), \ldots, \nu(f_d)].$$

Definition 4.3. We associate with the second-order smooth function F on $\mathbb{R}^d$, for some $d \geq 1$, the random functionals defined by

$$\begin{aligned} f \ &= \ (f_i)_{1 \leq i \leq d} \in \mathrm{Osc}(E)^d \\ &\mapsto F(m(X)(f)) = F(m(X)(f_1), \ldots, m(X)(f_d)) \in \mathbb{R}. \end{aligned} \tag{4.6}$$

Given a probability measure ν, and a collection of functions $(f_i)_{1 \leq i \leq d} \in \mathrm{Osc}(E)^d$, we set

$$D_\nu(F)(f) = \nabla F(\nu(f)) f^\top. \tag{4.7}$$

Notice that

$$\mathrm{osc}(D_\nu(F)(f)) \leq \|\nabla F(\nu(f))\|_1 := \sum_{i=1}^d \left| \frac{\partial F}{\partial u^i}(\nu(f)) \right|.$$

We also introduce the following constants

$$\|\nabla^2 F_f\|_1 := \sum_{i,j=1}^d \sup \left| \frac{\partial^2 F}{\partial u^i \partial u^j}(\nu(f)) \right|. \tag{4.8}$$

In the r.h.s. display, the supremum is taken over all probability measures $\nu \in \mathcal{P}(E)$.

The next theorem extends the exponential inequalities stated in theorem 4.1 to this class of nonlinear functionals. It also provides more precise concentration properties in terms of the variance functional σ defined in (4.1). The proof of this theorem is postponed to subsection 4.7.

Theorem 4.2. Let F be a second-order smooth function on $\mathbb{R}^d$, for some $d \geq 1$. For any collection of functions $(f_i)_{1 \leq i \leq d} \in \mathrm{Osc}(E)^d$, and any $N \geq 1$, the probability of the events

$$
\begin{aligned}
&[F(m(X)(f)) - F(\mu(f))] \\
&\quad \leq \frac{1}{2N} \|\nabla^2 F_f\|_1 [3/2 + \epsilon_0^{-1}(x)] \\
&\qquad + \|\nabla F(\mu(f))\|_1^{-1} \sigma^2(D_\mu(F)(f)) \epsilon_1^{-1} \left(\frac{x \|\nabla F(\mu(f))\|_1^2}{N \sigma^2(D_\mu(F)(f))} \right)
\end{aligned}
$$

is greater than $1 - e^{-x}$, for any $x \geq 0$. In the above display, $D_\mu(F)(f)$ stands for the first-order function defined in (4.7).

4.2.2 Empirical Processes

Our objective here is to extend the quantitative concentration theorems, Theorem 4.1 and Theorem 4.2, at the level of the empirical process associated with a class of function $\mathcal{F}$. These processes are given by the mapping

$$
f \in \mathcal{F} \mapsto V(X)(f) = \sqrt{N}(m(X) - \mu(f)).
$$

Our main result in this direction is the following theorem, whose proof is postponed to subsection 4.5.

Theorem 4.3. For any class of functions $\mathcal{F}$, with $I(\mathcal{F}) < \infty$, we have

$$
\pi_\psi(\|V(X)\|_{\mathcal{F}}) \leq 12^2 \int_0^2 \sqrt{\log(8 + \mathcal{N}(\mathcal{F}, \epsilon)^2)}\, d\epsilon.
$$

Remark 4.2. Using the fact that $\log(8+x^2) \leq 4\log x$, for any $x \geq 2$, we obtain the rather crude estimate

$$\int_0^2 \sqrt{\log(8+\mathcal{N}(\mathcal{F},\epsilon)^2)}\, d\epsilon \leq 2\int_0^2 \sqrt{\log \mathcal{N}(\mathcal{F},\epsilon)}\, d\epsilon.$$

We check the first observation, using the fact that $\theta(x) = 4\log x - \log(8+x^2)$ is a non decreasing function on $\mathbb{R}_+$, and $\theta(2) = \log(4\times 4) - \log(4\times 3) \geq 0$.

Various examples of classes of functions with finite covering and entropy integrals are given in the book by Van der Vaart and Wellner [93] (see, for instance, p. 86, p. 129, p. 135, and exercise 4 on p. 150, and p. 155). The estimation of the quantities introduced above often depends on several deep results on combinatorics, which are not discussed here.

To illustrate these mathematical objects, we mention that, for the set of indicator functions

$$\mathcal{F} = \{1_{\prod_{i=1}^d (-\infty, x_i]} \ ; \ (x_i)_{1\leq i\leq d} \in \mathbb{R}^d\} \tag{4.9}$$

of cells in $E = \mathbb{R}^d$, we have

$$\mathcal{N}(\varepsilon,\mathcal{F}) \leq c(d+1)(4e)^{d+1}\epsilon^{-2d}$$

for some universal constant $c < \infty$. This implies that

$$\sqrt{\log \mathcal{N}(\varepsilon,\mathcal{F})} \leq \sqrt{\log[c(d+1)(4e)^{d+1}]} + \sqrt{(2d)}\ \sqrt{\log(1/\epsilon)}.$$

An elementary calculation gives

$$\int_0^2 \sqrt{\log(1/\epsilon)} \leq 2\int_0^\infty x^2 e^{-x^2} dx = \sqrt{\pi/4} \leq 1$$

from which we conclude that

$$\int_0^2 \sqrt{\log \mathcal{N}(\mathcal{F},\epsilon)}\, d\epsilon \leq 2\sqrt{\log[c(d+1)(4e)^{d+1}]} + \sqrt{(2d)} \leq c'\sqrt{d} \tag{4.10}$$

for some universal constant $c < \infty$.

For $d = 1$, we also have that $\mathcal{N}(\varepsilon, \mathcal{F}) \leq 2/\epsilon^2$ (cf. p. 129 in [93]) and therefore

$$\int_0^2 \sqrt{\log \mathcal{N}(\mathcal{F}, \epsilon)}\, d\epsilon \leq 3\sqrt{2}.$$

Remark 4.3. In this section, we have assumed that the class of functions $\mathcal{F}$ is such that $\sup_{f \in \mathcal{F}} \|f\| \leq 1$. When $\sup_{f \in \mathcal{F}} \|f\| \leq c_{\mathcal{F}}$, for some finite constant $c_{\mathcal{F}}$, using Theorem 4.3, it is also readily checked that

$$\pi_\psi(\|V(X)\|_{\mathcal{F}}) \leq 12^2 \int_0^{2c_{\mathcal{F}}} \sqrt{\log\left(8 + \mathcal{N}(\mathcal{F}, \epsilon)^2\right)}\, d\epsilon. \tag{4.11}$$

We mention that the uniform entropy condition $I(\mathcal{F}) < \infty$ is required in Glivenko-Cantelli and Donsker theorems for empirical processes associated with non necessarily independent random sequences [36].

Arguing as above, we prove that the probability of the events

$$\|V(X)\|_{\mathcal{F}} \leq I_1(\mathcal{F}) \sqrt{x + \log 2}$$

is greater than $1 - e^{-x}$, for any $x \geq 0$, with some constant

$$I_1(\mathcal{F}) \leq 12^2 \int_0^2 \sqrt{\log\left(8 + \mathcal{N}(\mathcal{F}, \epsilon)^2\right)}\, d\epsilon.$$

As for the marginal models 4.6, our next objective is to extend Theorem 4.5 to the empirical processes associated with some classes of functions. Here, we consider the empirical processes

$$f \in \mathcal{F}_i \mapsto m(X)(f) \in \mathbb{R}$$

associated with d classes of functions $\mathcal{F}_i$, $1 \leq i \leq d$, defined in subsection 4.1. We further assume that $\|f_i\| \vee \mathrm{osc}(f_i) \leq 1$, for any $f_i \in \mathcal{F}_i$, and we set

$$\mathcal{F} := \prod_{1 \leq i \leq d} \mathcal{F}_i \quad \text{and} \quad \pi_\psi(\|V(X)\|_{\mathcal{F}}) := \sup_{1 \leq i \leq d} \pi_\psi(\|V(X)\|_{\mathcal{F}_i}).$$

Using Theorem 4.3, we mention that

$$\pi_\psi(\|V(X)\|_{\mathcal{F}}) \leq 12^2 \int_0^2 \sqrt{\log\left(8 + \mathcal{N}(\mathcal{F}, \epsilon)^2\right)}\, d\epsilon$$

with

$$\mathcal{N}(\mathcal{F},\epsilon) := \sup_{1\le i\le d} \mathcal{N}(\mathcal{F}_i,\epsilon).$$

We set

$$\|\nabla F_\mu\|_\infty := \sup\left|\frac{\partial F}{\partial u^i}(\mu(f))\right| \quad \text{and} \quad \|\nabla^2 F\|_\infty = \sup\left|\frac{\partial^2 F}{\partial u^i \partial u^j}(\nu(f))\right|.$$

The supremum in the l.h.s. is taken over all $1 \le i \le d$ and all $f \in \mathcal{F}$; and the supremum in the r.h.s. is taken over all $1 \le i,j \le d$, $\nu \in \mathcal{P}(E)$, and all $f \in \mathcal{F}$.

We are now in a position to state the final main result of this subsection. The proof of the next theorem is postponed to the end of subsection 4.7.

Theorem 4.4. Let F be a second-order smooth function on $\mathbb{R}^d$, for some $d \ge 1$. For any classes of functions $\mathcal{F}_i$, $1 \le i \le d$, and for any $x \ge 0$, the probability of the following events

$$\begin{aligned}&\sup_{f\in\mathcal{F}} |F(m(X)(f)) - F(\mu(f))| \\ &\qquad \le \frac{d}{\sqrt{N}}\pi_\psi(\|V(X)\|_{\mathcal{F}})\|\nabla F_\mu\|_\infty(1+2\sqrt{x}) \\ &\qquad\quad + \frac{1}{2N}\|\nabla^2 F\|_\infty(d\ \pi_\psi(\|V(X)\|_{\mathcal{F}}))^2\left(1+\epsilon_0^{-1}\left(\frac{x}{2}\right)\right)\end{aligned}$$

is greater than $1 - e^{-x}$.

4.3 A Reminder on Orlicz' Norms

Here, we have collected some important properties of Orlicz' norms. Subsection 4.3.1 is concerned with rather elementary comparison properties. In subsection 4.3.2, we present a natural way to obtain Laplace estimates, and related concentration inequalities, using simple Orlicz' norm upper bounds.

4.3.1 Comparison Properties

This short subsection is mainly concerned with the proof of the following three comparison properties.

Lemma 4.5. For any nonnegative variables (Y_1, Y_2) we have

$$Y_1 \leq Y_2 \Longrightarrow \pi_\psi(Y_1) \leq \pi_\psi(Y_2)$$

as well as

$$(\forall\, m \geq 0 \mathbb{E}(Y_1^{2m}) \leq \mathbb{E}(Y_2^{2m})) \Rightarrow \pi_\psi(Y_1) \leq \pi_\psi(Y_2). \tag{4.12}$$

In addition, for any pair of independent random variables (X, Y) on some measurable state space, and any measurable function f, we have

$$(\pi_\psi(f(x, Y)) \leq c \text{ for } \mathbb{P}\text{-a.e. x}) \Longrightarrow \pi_\psi(f(X, Y)) \leq c. \tag{4.13}$$

Proof:
The first assertion is immediate, and the second assertion comes from the fact that

$$\mathbb{E}\left(\exp\left(\frac{Y_1}{\pi_\psi(Y_2)}\right)^2 - 1\right) \leq \sum_{m \geq 1} \frac{1}{m!}\, \frac{\mathbb{E}(Y_2^{2m})}{\pi_\psi(Y_2)^{2m}} = \mathbb{E}\left(\frac{Y_2}{\pi_\psi(Y_2)}\right) \leq 1.$$

The last assertion comes from the fact that

$$\mathbb{E}(\mathbb{E}(\psi(f(X, Y)/c)|X)) \leq 1 \Rightarrow \pi_\psi(f(X, Y)) \leq c.$$

This ends the proof of the lemma. ∎

4.3.2 Concentration Properties

The following lemma provides a simple way to transfer a control on Orlicz' norm into moment or Laplace estimates, which in turn can be used to derive quantitative concentration inequalities .

Lemma 4.6. For any non negative random variable Y, and any integer $m \geq 0$, we have

$$\mathbb{E}(Y^{2m}) \leq m!\ \pi_\psi(Y)^{2m} \quad \text{and} \quad \mathbb{E}(Y^{2m+1}) \leq (m+1)!\pi_\psi(Y)^{2m+1}. \tag{4.14}$$

In addition, for any $t \geq 0$ we have the Laplace estimates

$$\mathbb{E}(e^{tY}) \leq \min\left(2e^{\frac{1}{4}(t\pi_\psi(Y))^2}, (1 + t\pi_\psi(Y))e^{(t\pi_\psi(Y))^2}\right).$$

In particular, for any $x \geq 0$ the probability of the event

$$Y \leq \pi_\psi(Y)\sqrt{x + \log 2} \tag{4.15}$$

is greater than $1 - e^{-x}$.

Remark 4.4. For a Gaussian and centered random variable Y, s.t. $E(Y^2) = 1$, we recall that $\pi_\psi(Y) = \sqrt{8/3}$. In this situation, letting $y = \sqrt{8(x + \log 2)/3}$ in (4.15), we find that

$$\mathbb{P}(|Y| \geq y) \leq 2e^{-\frac{1}{2}\frac{3}{4}y^2}.$$

Working directly with the Laplace Gaussian function $\mathbb{E}(e^{tY}) = e^{t^2/2}$, we remove the factor 3/4. In this sense, we loose a factor 3/4 using the Orlicz's concentration property (4.15).

In this situation, the l.h.s. moment estimate in (4.14) takes the form

$$b(2m)^{2m} = \frac{(2m)!}{m!}\, 2^{-m} \leq m!\, (8/3)^m \tag{4.16}$$

while, using Stirling's approximation of the factorials, we obtain the estimate

$$\frac{(2m)!}{m!^2} \simeq \sqrt{2/m}\, 4^m (\leq (8/3)^m).$$

Remark 4.5. Given a sequence of independent Gaussian and centered random variables Y_i, s.t. $E(Y_i^2) = 1$, for $i \geq 1$, and any sequence of non negative numbers a_i, we have

$$\pi_\psi\left(\sum_{i=1}^n a_i Y_i\right) = \sqrt{8/3}\sqrt{\sum_{i=1}^n a_i^2} := \sqrt{8/3}\|a\|_2$$

while

$$\sum_{i=1}^n a_i \pi_\psi(Y_i) = \sqrt{8/3}\sum_{i=1}^n a_i := \sqrt{8/3}\|a\|_1.$$

Notice that

$$\|a\|_2 \leq \|a\|_1 \leq \sqrt{n}\, \|a\|_2.$$

When the coefficients a_i are almost equal, we can lose a factor $\sqrt{n}$ using the triangle inequality, instead of estimating directly with the Orlicz norm of the Gaussian mixture. In this sense, it is always preferable to avoid the use of the triangle inequality, and to estimate directly the Orlicz norms of linear combinations of "almost Gaussian" random variables.

Now, we come to the proof of the lemma.

Proof of Lemma 4.6:
For any $m \geq 1$, we have

$$x^{2m} \leq m! \sum_{n \geq 1} \frac{x^{2n}}{n!} = m!\ \psi(x)$$

$$\Downarrow$$

$$\mathbb{E}\left(\left[\frac{Y}{\pi_\psi(Y)}\right]^{2m}\right) \leq m! \mathbb{E}\left(\psi\left(\frac{Y}{\pi_\psi(Y)}\right)\right) \leq m!.$$

For odd integers, we simply use the Cauchy-Schwarz inequality to check that

$$\mathbb{E}(Y^{2m+1})^2 \leq \mathbb{E}(Y^{2m})\mathbb{E}(Y^{2(m+1)}) \leq (m+1)!^2\ \pi_\psi(Y)^{2(2m+1)}.$$

This ends the proof of the first assertion.

We use Cauchy-Schwarz's inequality to check that

$$\mathbb{E}(Y^{2m+1})^2 \leq \mathbb{E}(Y^{2m})\ \mathbb{E}(Y^{2(m+1)}) \leq (m+1)!^2\ \pi_\psi(Y)^{2(2m+1)}$$

for any non negative random variable Y, so that

$$\mathbb{E}(Y^{2m+1}) \leq (m+1)!\ \pi_\psi(Y)^{(2m+1)}.$$

Recalling that $(2m)! \geq m!^2$ and $(m+1) \leq (2m+1)$, we find that

$$\begin{aligned}
\mathbb{E}(e^{tY}) &= \sum_{m \geq 0} \frac{t^{2m}}{(2m)!}\mathbb{E}(Y^{2m}) + \sum_{m \geq 0} \frac{t^{2m+1}}{(2m+1)!}\ \mathbb{E}(Y^{2m+1}) \\
&\leq \sum_{m \geq 0} \frac{t^{2m}}{m!}\ \pi_\psi(Y)^{2m} + \sum_{m \geq 0} \frac{t^{2m+1}}{m!}\ \pi_\psi(Y)^{(2m+1)} \\
&= (1 + t\pi_\psi(Y)) \exp\left(t\pi_\psi(Y)\right)^2.
\end{aligned}$$

On the other hand, using the estimate

$$tY = \left(\frac{t\pi_\psi(Y)}{\sqrt{2}}\right)\left(\frac{\sqrt{2}\ Y}{\pi_\psi(Y)}\right) \leq \frac{(t\pi_\psi(Y))^2}{4} + \left(\frac{Y}{\pi_\psi(Y)}\right)^2$$

we prove that

$$\mathbb{E}(e^{tY}) \leq 2\ \exp\left(\frac{(t\pi_\psi(Y))^2}{4}\right).$$

The end of the proof of the Laplace estimates is now completed. To prove the last assertion, we use the fact that for any $y \geq 0$

$$\begin{aligned}\mathbb{P}(Y \geq y) &\leq 2\exp\Big(-\sup_{t\geq 0}(ty - (t\pi_\psi(Y))^2/4)\Big)\\ &= 2\exp\big[-(y/\pi_\psi(Y))^2\big].\end{aligned}$$

This implies that

$$\mathbb{P}(Y \geq \pi_\psi(Y)\sqrt{x+\log 2}) \leq 2\exp\left[-(x+\log 2)\right] = e^{-x}.$$

This ends the proof of the lemma. ∎

4.3.3 Maximal Inequalities

Let us now assemble the Orlicz's norm properties derived in subsection 4.3 to establish a series of more or less well known maximal inequalities. More general results can be found in the books [81, 93], or in the lecture notes [56].

We emphasize that in the literature on empirical processes, maximal inequalities are often presented in terms of a universal constant c without further information on their magnitude. Now, we shall try to estimate explicitly some of these universal constants.

To begin with, we consider a couple of maximal inequalities over finite sets.

Lemma 4.7. For any finite collection of nonnegative random variables $(Y_i)_{i\in I}$, and any collection of nonnegative numbers $(a_i)_{i\in I}$, we have

$$\sup_{i\in I}\mathbb{E}(\psi(Y_i/a_i)) \leq 1 \Rightarrow \mathbb{E}\Big(\max_{i\in I} Y_i\Big) \leq \psi^{-1}(|I|) \times \max_{i\in I} a_i.$$

Proof:
We check this claim using the following estimates

$$\begin{aligned}\psi\left(\frac{\mathbb{E}(\max_{i\in I} Y_i)}{\max_{i\in I} a_i}\right) &\leq \psi\Big(\mathbb{E}\Big(\max_{i\in I}(Y_i/a_i)\Big)\Big)\\ &\leq \mathbb{E}\Big(\psi\Big(\max_{i\in I}(Y_i/a_i)\Big)\Big)\\ &\leq \mathbb{E}\left(\sum_{i\in I}\psi((Y_i/a_i))\right) \leq |I|.\end{aligned}$$

This ends the proof of the lemma. ∎

Proceeding further, we prove the following lemma.

Lemma 4.8. For any finite collection of non negative random variables $(Y_i)_{i\in I}$, we have

$$\pi_\psi\Big(\max_{i\in I} Y_i\Big) \leq \sqrt{6\log(8+|I|)}\ \max_{i\in I}\pi_\psi(Y_i).$$

Proof:
Without lost of generality, we assume that $\max_{i\in I}\pi_\psi(Y_i) \leq 1$, and $I = \{1,\ldots,|I|\}$. In this situation, it suffices to check that

$$\psi\left(\frac{\max_{1\leq i\leq |I|} Y_i}{\sqrt{6\log(8+|I|)}}\right) \leq \psi\left(\max_{1\leq i\leq |I|}\frac{Y_i}{\sqrt{6\log(8+i)}}\right) \leq 1.$$

First, we notice that for any $i \geq 1$ and $x \geq 3/2$ we have

$$\frac{1}{\log(8+i)} + \frac{1}{\log x} \leq \frac{1}{\log 9} + \frac{1}{\log(3/2)} \leq 3$$

and therefore

$$3\log(8+i)\log(x) \geq \log(x(8+i)).$$

We check the first estimate using the fact that

$$\begin{aligned}\log(3) &\leq 5\log(3/2) \Rightarrow \log(3)+\log(3/2)\\ &\leq 6\log(3/2) \leq 3\log(3/2)\log(9).\end{aligned}$$

Using these observations, we have

$$\begin{aligned}
&\mathbb{P}\left(\max_{1\le i\le |I|}\left(\frac{Y_i}{\sqrt{6\log(8+i)}}\right)^2 > \log x\right)\\
&\quad = \mathbb{P}\left(\max_{1\le i\le |I|}\left(\frac{Y_i}{\sqrt{6\log(x)\log(8+i)}}\right)^2 > 1\right)\\
&\quad \le \mathbb{P}\left(\max_{1\le i\le |I|}\frac{Y_i}{\sqrt{2\log(x(8+i))}} > 1\right)\\
&\quad \le \sum_{i=1}^{|I|}\mathbb{P}(Y_i > \sqrt{2\log(x(8+i))}) \le \sum_{i=1}^{|I|} e^{-2\log(x(8+i))}\mathbb{E}(e^{Y_i^2}).
\end{aligned}$$

This implies that

$$\begin{aligned}
&\mathbb{P}\left(\max_{1\le i\le |I|}\left(\frac{Y_i}{\sqrt{6\log(8+i)}}\right)^2 > \log x\right)\\
&\qquad \le \frac{2}{x^2}\sum_{i=1}^{|I|}\frac{1}{(8+i)^2} \le \frac{2}{x^2}\int_8^\infty \frac{1}{u^2}\,du = \frac{1}{(2x)^2}.
\end{aligned}$$

If we set

$$Z_I := \exp\left\{\left(\max_{1\le i\le |I|}\frac{Y_i}{\sqrt{6\log(8+i)}}\right)^2\right\}$$

then we have

$$\begin{aligned}
\mathbb{E}(Z_I) &= \int_0^\infty \mathbb{P}(Z_I > x)dx\\
&\le \frac{3}{2} + \int_{\frac{3}{2}}^\infty \frac{1}{(2x)^2}dx = \frac{3}{2}\left(1+\frac{1}{4}\right) = \frac{15}{8} \le 2
\end{aligned}$$

and therefore

$$\psi\left(\max_{1\le i\le |I|}\left(\frac{Y_i}{\sqrt{6\log(8+i)}}\right)\right) \le 1.$$

This ends the proof of the lemma. ∎

The following technical lemma is pivotal in the analysis of maximal inequalities for sequences of random variables indexed by infinite but separable subsets equipped with a pseudo-metric, under some Lipschitz regularity conditions w.r.t. the Orlicz's norm.

Lemma 4.9. We assume that the index set (I, d) is a separable, and totally bounded pseudo-metric space, with finite diameter

$$d(I) := \sup_{(i,i)\in I^2} d(i,j) < \infty.$$

We let $(Y_i)_{i\in I}$ be a separable and $\mathbb{R}$-valued stochastic process indexed by I and such that

$$\pi_\psi(Y_i - Y_j) \leq c\ d(i,j)$$

for some finite constant $c < \infty$. We also assume that $Y_{i_0} = 0$, for some $i_0 \in I$. Then, we have

$$\pi_\psi(\sup_{i\in I} Y_i) \leq 12c \int_0^{d(I)} \sqrt{6 \log\left(8 + \mathcal{N}(I,d,\epsilon)^2\right)}\ d\epsilon.$$

Proof:
Replacing Y_i by $Y_i/d(I)$, and d by $d/d(I)$, there is no loss of generality to assume that $d(I) \leq 1$. In the same way, replacing Y_i by Y_i/c, we can also assume that $c \leq 1$. For a given finite subset $J \subset I$, with $i_0 \in J$, we let $J_k = \{i_1^k, \ldots, i_{n_k}^k\} \subset J$, be the centers of $n_k = \mathcal{N}(J,d,2^{-k})$ balls of radius at most 2^{-k} covering J. For $k = 0$, we set $J_0 = \{i_0\}$. We also consider the mapping $\theta_k\ :\ i \in J \mapsto \theta_k(i) \in J_k$ s.t.

$$\sup_{i\in J} d(\theta_k(i), i) \leq 2^{-k}.$$

The set J being finite, there exists some integer $k_J^\star$ s.t. $d(\theta_k(i), i) = 0$, for any $k \geq k_J^\star$; and therefore $Y_i = Y_{\theta_k(i)}$, for any $i \in J$, and any $k \geq k_J^\star$. This implies that

$$Y_i = \sum_{k=1}^{k_J^\star} [Y_{\theta_k(i)} - Y_{\theta_{k-1}(i)}].$$

We also notice that

$$d(\theta_k(i),\theta_{k-1}(i)) \leq d(\theta_k(i),i) + d(i,\theta_{k-1}(i)) \leq 2^{-k} + 2^{-(k-1)} = 3 \times 2^{-k}$$

and

$$\sup_{(i,j)\in(J_k\times J_{k-1})\ :\ d(i,j)\leq 3\times 2^{-k}} \pi_\psi(Y_i - Y_j) \leq 3 \times 2^{-k}.$$

Using Lemma 4.8 we prove that

$$\begin{aligned}\pi_\psi(\sup_{i\in J} Y_i) &\leq \sum_{k=1}^{k_J^\star} \pi_\psi(\sup_{i\in J}[Y_{\theta_k(i)} - Y_{\theta_{k-1}(i)})] \\ &\leq 3\sum_{k=1}^{k_J^\star} \sqrt{6\log\left(8+\mathcal{N}(J,d,2^{-k})^2\right)}\ 2^{-k}.\end{aligned}$$

On the other hand, we have

$$2(2^{-k} - 2^{-(k+1)}) = 2^{-k}$$

and

$$\sqrt{6\log\left(8+\mathcal{N}(J,d,2^{-k})^2\right)}2^{-k} \leq 2\int_{2^{-(k+1)}}^{2^{-k}} \sqrt{6\log\left(8+\mathcal{N}(J,d,\epsilon)^2\right)}d\epsilon$$

from which we conclude that

$$\pi_\psi(\sup_{i\in J} Y_i) \leq 6\int_0^{1/2} \sqrt{6\log\left(8+\mathcal{N}(J,d,\epsilon)^2\right)}d\epsilon.$$

Using the fact that the ϵ-balls with center in I and intersecting J are necessarily contained in an (2ϵ)-ball with center in J, we also have

$$\mathcal{N}(J,d,2\epsilon) \leq \mathcal{N}(I,d,\epsilon).$$

This implies that

$$\pi_\psi(\sup_{i\in J} Y_i) \leq 12\int_0^{1} \sqrt{6\log\left(8+\mathcal{N}(I,d,\epsilon)^2\right)}d\epsilon.$$

The end of the proof is now a direct consequence of the monotone convergence theorem with increasing series of finite subsets exhausting I. This ends the proof of the lemma. ∎

4.4 Marginal Inequalities

This subsection is mainly concerned with the proof of the Theorem 4.1. This result is a more or less direct consequence of the following technical lemma, which is of separate interest.

Lemma 4.10. 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

$$\mathbb{E}(|M_n|^m)^{\frac{1}{m}} \leq \; b(m)\mathbb{E}([M]_n^{m'/2})^{\frac{1}{m'}} \tag{4.17}$$

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 (1.27). In addition, for any $m \geq 2$, we have

$$\mathbb{E}(|M_n|^m)^{\frac{1}{m}} \leq b(m)\sqrt{(n+1)} \left(\frac{1}{n+1} \sum_{0 \leq p \leq n} \mathbb{E}(|\Delta_p|^{m'}) \right)^{\frac{1}{m'}}. \tag{4.18}$$

Proof of Theorem 4.1: We consider a collection of independent copies $X' = (X'^i)_{i \geq 1}$ of the random variables $X = (X^i)_{i \geq 1}$. We consider the martingale sequence $M = (M_i)_{1 \leq i \leq N}$ with symmetric and independent increments defined for any $1 \leq j \leq N$ by the following formula

$$M_j := \frac{1}{\sqrt{N}} \sum_{i=1}^{j} [f(X^i) - f(X'^i)].$$

By construction, we have

$$V(X)(f) \; = \; \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \; (f(X^i) - \mu^i(f)) = \mathbb{E}(M_N | X).$$

Combining this conditioning property with the estimates provided in Lemma 4.10, the proof of the first assertion is now easily completed.

The Orlicz norm estimate (4.4) comes from the fact that for any $f \in \mathrm{Osc(E)}$, we have

$$\mathbb{E}(|V(X)(f)|^{2m}) \leq b(2m)^{2m} = \mathbb{E}(U^{2m})$$

for a Gaussian and centered random variable U, s.t. $E(U^2) = 1$. Using the comparison lemma, lemma 4.12, we find that

$$\pi_\psi(V(X)(f)) \leq \pi_\psi(U) = \sqrt{8/3}.$$

Applying Kintchine's inequalities (4.17), we prove that

$$\mathbb{E}(|V(X)(f)|^m)^{\frac{1}{m}} \leq b(m)\mathbb{E}\left(\left[\frac{1}{N}\sum_{i=1}^N [f(X^j) - f(X'^j)]^2\right]^{m'/2}\right)^{1/m'}.$$

By construction, we notice that for any $f \in \mathrm{osc(E)}$, and any $p \geq 2$, we have

$$\frac{1}{N}\sum_{j=1}^N \mathbb{E}([f(X^j) - f(X'^j)]^p) \leq 2\sigma(f)^2.$$

By the Rosenthal type inequality stated in Theorem 2.5 in [65], for any sequence of nonnegative, independent and bounded random variables $(Y_i)_{i\geq 1}$, we have the rough estimate

$$\mathbb{E}\left[\sum_{i=1}^N Y_i^p\right]^{1/p} \leq 2p \max\left(\sum_{i=1}^N \mathbb{E}(Y_i), \left[\sum_{i=1}^N \mathbb{E}(Y_i^p)\right]^{1/p}\right)$$

for any $p \geq 1$. If we take $p = m'/2$, and

$$Y_i = \frac{1}{N}[f(X^i) - f(X'^i)]^2$$

we prove that

$$\mathbb{E}\left(\left[\frac{1}{N}\sum_{i=1}^N [f(X^i) - f(X'^i)]^2\right]^{m'/2}\right)^{2/m'}$$
$$\leq 4m \max\left(2\sigma(f)^2, \frac{1}{N^{1-\frac{2}{m'}}}\left[2\sigma(f)^2\right]^{2/m'}\right)$$

for any $f \in \text{osc}(\text{E})$. Using Stirling's approximation of factorials

$$\sqrt{2\pi n} n^n e^{-n} \leq n! \leq e\sqrt{2\pi n} n^n e^{-n}$$

for any $p \geq 1$ we have

$$(2p)^p / b(2p)^{2p} = 2^{2p} p^p p! / (2p)! \leq e^{p+1} \leq 3^{2p}$$

and

$$\begin{aligned}(2p+1)^{p+1/2} / b(2p+1)^{2p+1} &= (2p+1)^{p+1} 2^p p! / (2p+1)! \\ &\leq e^{p+2} \leq 3^{2p+1}.\end{aligned}$$

This implies that

$$m^{m/2} / b(m)^m \leq 3^m \Rightarrow \sqrt{m}\ b(m) \leq 3 b(m)^2$$

for any $m \geq 1$. This ends the proof of the theorem. ∎

Now, we come to the proof of the lemma.

Proof of Lemma 4.10:
We prove the lemma by induction on the parameter n. The result is clearly satisfied for $n = 0$. Suppose the estimate (4.17) 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 Δ_n are null. Consequently, we find that

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

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

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

To take the final step, we use the fact that

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

and $(2p)_p \geq 2^p$, to conclude that

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

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

$$\mathbb{E}(|M_n|^{2m+1})^2 \leq \mathbb{E}(M_n^{2m})\mathbb{E}(M_n^{2(m+1)})$$
$$\leq 2^{-(2m+1)}(2m)_m(2(m+1))_{(m+1)}\mathbb{E}([M]_n^{m+1})^{\frac{2m+1}{m+1}}.$$

We conclude that

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

The proof of (4.17) is now completed. Now, we come to the proof of (4.18). For any $m' \geq 2$ we have

$$\left[\frac{1}{n+1}\sum_{0\leq p\leq n}\Delta_p^2\right]^{m'/2} \leq \frac{1}{n+1}\sum_{0\leq p\leq n}\mathbb{E}(|\Delta_p|^{m'})$$

and therefore

$$\mathbb{E}([M]_n^{m'/2})^{\frac{1}{m'}} \leq (n+1)^{1/2}\left(\frac{1}{n+1}\sum_{0\leq p\leq n}\mathbb{E}(|\Delta_p|^{m'})\right)^{\frac{1}{m'}}.$$

This ends the proof of the lemma. ∎

4.5 Maximal Inequalities

To prove Theorem 4.3, we begin with the basic symmetrization technique. We consider a collection of independent copies $X' = (X'^i)_{i\geq 1}$ of the random variables $X = (X^i)_{i\geq 1}$. Let $\varepsilon = (\varepsilon_i)_{i\geq 1}$ constitute a sequence that is independent and identically distributed with

$$P(\varepsilon_1 = +1) = P(\varepsilon_1 = -1) = 1/2.$$

We also consider the empirical random field sequences

$$V_\epsilon(X) := \sqrt{N} m_\varepsilon(X).$$

We also assume that (ϵ, X, X') are independent. We associate with the pairs (ϵ, X) and (ϵ, X') the random measures $m_\epsilon(X) = \frac{1}{N}\sum_{i=1}^N \epsilon_i\ \delta_{X^i}$ and $m_\epsilon(X') = \frac{1}{N}\sum_{i=1}^N \epsilon_i\ \delta_{X'^i}$.

We notice that

$$\begin{aligned}\|m(X) - \mu\|_{\mathcal{F}}^p &= \sup_{f\in\mathcal{F}} |m(X)(f) - \mathbb{E}(m(X')(f))|^p \\ &\leq \mathbb{E}(\|m(X) - m(X')\|_{\mathcal{F}}^p\ |X)\end{aligned}$$

and, in view of the symmetry of the random variables $(f(X^i) - f(X'^i))_{i\geq 1}$ we have

$$\mathbb{E}(\|m(X) - m(X')\|_{\mathcal{F}}^p) = \mathbb{E}(\|m_\epsilon(X) - m_\epsilon(X')\|_{\mathcal{F}}^p)$$

from which we conclude that

$$E(\|V(X)\|_{\mathcal{F}}^p) \leq 2^p E(\|V_\epsilon(X)\|_{\mathcal{F}}^p). \tag{4.19}$$

By using the Chernov-Hoeffding inequality for any $x^1, \ldots, x^N \in E$, the empirical process

$$f \longrightarrow V_\epsilon(x)(f) := \sqrt{N} m_\varepsilon(x)(f)$$

is sub-Gaussian for the norm $\|f\|_{L_2(m(x))} = m(x)(f^2)^{1/2}$. Namely, for any couple of functions f, g and any $\delta > 0$ we have

$$\mathbb{E}([V_\epsilon(x)(f) - V_\epsilon(x)(g)]^2) = \|f - g\|_{\mathbb{L}_2(m(x))}^2$$

and by Hoeffding's inequality

$$P(|V_\epsilon(x)(f) - V_\epsilon(x)(g)| \geq \delta) \leq 2e^{-\frac{1}{2}\delta^2/\|f-g\|_{\mathbb{L}_2(m(x))}^2}.$$

If we set $Z = (\frac{V_\epsilon(x)(f)}{\sqrt{6}\|f\|_{\mathbb{L}_2(m(x))}})^2$, then we find that

$$\begin{aligned}\mathbb{E}(e^Z) - 1 &= \int_0^\infty e^t \mathbb{P}(Z \geq t) dt \\ &= \int_0^\infty e^t \mathbb{P}(|V_\epsilon(x)(f)| \geq \sqrt{6t}\|f\|_{\mathbb{L}_2(m(x))}) dt \\ &\leq 2\int_0^\infty e^t e^{-3t} dt = 1\end{aligned}$$

from which we conclude that

$$\pi_\psi(V_\epsilon(x)(f) - V_\epsilon(x)(g)) \leq \sqrt{6}\|f - g\|_{\mathbb{L}_2(m(x))}.$$

Combining the maximal inequalities stated in Lemma 4.9 and the conditioning property (4.13) we find that

$$\pi_\psi(\|V_\epsilon(X)\|_{\mathcal{F}}) \leq J(\mathcal{F})$$

with

$$J(\mathcal{F}) \leq 26^2 \int_0^2 \sqrt{\log\left(8 + \mathcal{N}(\mathcal{F}, \epsilon)^2\right)}\, d\epsilon \leq cI(\mathcal{F}) < \infty$$

for some finite universal constant $c < \infty$. Combining (4.19) with (4.12), this implies that

$$\pi_\psi(\|V(X)\|_{\mathcal{F}}) \leq 2\ J(\mathcal{F}).$$

This ends the proof of the theorem. ∎

4.6 Cramér-Chernov Inequalities

4.6.1 Some Preliminary Convex Analysis

Here, we present some basic Cramér-Chernov tools to derive quantitative concentration inequalities. We begin by recalling some preliminary convex analysis on Legendre-Fenchel transforms. We associate with any convex function

$$L : t \in \mathrm{Dom(L)} \mapsto L(t) \in \mathbb{R}_+$$

defined in some domain $\mathrm{Dom(L)} \subset \mathbb{R}_+$, with $L(0) = 0$, the Legendre-Fenchel transform $L^\star$ defined by the variational formula

$$\forall\, \lambda \geq 0 \quad L^\star(\lambda) := \sup_{t \in \mathrm{Dom(L)}} (\lambda t - L(t)).$$

Note that $L^\star$ is a convex increasing function with $L^\star(0) = 0$ and its inverse $(L^\star)^{-1}$ is a concave increasing function.

We let L_A be the log-Laplace transform of a random variable A defined on some domain $\mathrm{Dom}(L_A) \subset \mathbb{R}_+$ by the formula

$$L_A(t) := \log \mathbb{E}(e^{tA}).$$

Hölder's inequality implies that L_A is convex. Using the Cramér-Chernov-Chebychev inequality, we find that

$$\log \mathbb{P}(A \geq \lambda) \leq -L_A^\star(\lambda) \quad \text{and} \quad \mathbb{P}(A \geq (L_A^\star)^{-1}(x)) \leq e^{-x}$$

for any $\lambda \geq 0$ and any $x \geq 0$.

The next lemma provides some key properties of Legendre-Fenchel transforms, which will be used in several places to further the development of this monograph.

Lemma 4.11.

- For any convex functions (L_1, L_2), such that

  $$\forall t \in \mathrm{Dom}(L_2) \quad L_1(t) \leq L_2(t) \quad \text{and} \quad \mathrm{Dom}(L_2) \subset \mathrm{Dom}(L_1)$$

  we have

  $$L_2^\star \leq L_1^\star \quad \text{and} \quad (L_1^\star)^{-1} \leq (L_2^\star)^{-1}$$

- if we have

  $$\forall t \in v^{-1}\mathrm{Dom}(L_2) = \mathrm{Dom}(L_1) \quad L_1(t) = u\ L_2(vt)$$

  for some positive numbers $(u, v) \in \mathbb{R}_+^2$, then we have

  $$L_1^\star(\lambda) = u\ L_2^\star\left(\frac{\lambda}{uv}\right) \quad \text{and} \quad (L_1^\star)^{-1}(x) = uv\ (L_2^\star)^{-1}\left(\frac{x}{u}\right)$$

  for any $\lambda \geq 0$, and any $\forall x \geq 0$.
- Let A be a random variable with a finite log-Laplace transform. For any $a \in \mathbb{R}$, we have

  $$L_A(t) = -at + L_{A+a}(t)$$

  as well as

  $$L_A^\star(\lambda) = L_{A+a}^\star(\lambda + a) \quad \text{and} \quad (L_A^\star)^{-1}(x) = -a + (L_{A+a}^\star)^{-1}(x).$$

We illustrate this technical lemma with the detailed analysis of three convex increasing functions, which will be of use in the further development of this monograph

- $L(t) = t^2/(1-t)$, $t \in [0,1]$
- $L_0(t) := -t - \frac{1}{2}\log(1-2t)$, $t \in [0,1/2]$.
- $L_1(t) := e^t - 1 - t$.

In the first situation, we readily check that

$$L'(t) = \frac{1}{(1-t)^2} - 1 \quad \text{and} \quad L''(t) = \frac{2}{(1-t)^3}.$$

An elementary manipulation yields that

$$L^\star(\lambda) = (\sqrt{\lambda+1} - 1)^2$$

and

$$(L^\star)^{-1}(x) = (1+\sqrt{x})^2 - 1 = x + 2\sqrt{x}.$$

In the second situation, we have

$$L_0'(t) = \frac{1}{1-2t} - 1 \quad \text{and} \quad L_0''(t) = \frac{2}{(1-2t)^2}$$

from which we find that

$$L_0^\star(\lambda) = \frac{1}{2}(\lambda - \log(1+\lambda)).$$

We also notice that

$$L_0(t) = t^2 \sum_{p\geq 0} \frac{2}{2+p}\ (2t)^p \leq \overline{L}_0(t) := \frac{t^2}{1-2t} = \frac{1}{4}\ L(2t)$$

for every $t \in [0, 1/2[$ Using Lemma 4.11, we prove that

$$\begin{aligned} \overline{L}_0^\star(\lambda) &= \frac{1}{4}\ L^\star(2\lambda) \leq L_0^\star(\lambda) \\ (L_0^\star)^{-1}(x) &\leq (\overline{L}_0^\star)^{-1}(x) = \frac{1}{2}(L^\star)^{-1}(4x) = 2(x+\sqrt{x}). \end{aligned} \tag{4.20}$$

In the third situation, we have

$$L_1'(t) = e^t - 1 \quad \text{and} \quad L_1''(t) = e^t$$

from which we conclude that

$$L_1^\star(\lambda) = (1+\lambda)\log(1+\lambda) - \lambda.$$

On the other hand, using the fact that $2 \times 3^p \leq (p+2)!$, for any $p \geq 0$, we prove that we have

$$L_1(t) = \frac{t^2}{2} \sum_{p\geq 0} \frac{2\times 3^p}{(p+2)!} \left(\frac{t}{3}\right)^p \leq \overline{L}_1(t) := \frac{t^2}{2(1-t/3)} = \frac{9}{2} L\left(\frac{t}{3}\right)$$

for every $t \in [0, 1/3]$. This implies that

$$\overline{L}_1^\star(\lambda) = \frac{9}{2}\ L^\star(\frac{2\lambda}{3}) \leq L_1^\star(\lambda)$$

and therefore

$$(L_1^\star)^{-1}(x) \leq (\overline{L}_1^\star)^{-1}(x) = \frac{3}{2}(L^\star)^{-1}\left(\frac{2x}{9}\right) = \left(\frac{x}{3} + \sqrt{2x}\right). \tag{4.21}$$

Another crucial ingredient in the concentration analysis of the sum of two random variables is the deep technical lemma of Bretagnolle and Rio [84]. In the further development of this section, we use this argument to obtain a large family of concentration inequalities that are asymptotically "almost sharp" in a wide variety of situations.

Lemma 4.12 (Bretagnolle & Rio [84]). For any pair of random variables A and B with a finite log-Laplace transform in the neighborhood of 0, we have

$$\forall x \geq 0 \quad (L_{A+B}^\star)^{-1}(x) \leq (L_A^\star)^{-1}(x) + (L_B^\star)^{-1}(x). \tag{4.22}$$

We also quote the following reverse type formulae that allows us to turn most of the concentration inequalities developed in this monograph into Bernstein-style exponential inequalities.

Lemma 4.13. For any $(u, v) \in \mathbb{R}_+$, we have

$$u(L_0^\star)^{-1}(x) + v\ (L_1^\star)^{-1}(x) \leq (L_{a(u,v),b(u,v)}^\star)^{-1}(x)$$

with the functions

$$a(u, v) := \left(2u + \frac{v}{3}\right) \quad \text{and} \quad b(u, v) := (\sqrt{2}\ u + v)^2$$

and the Laplace function

$$L_{a,b}(t) = \frac{b}{2a^2}\ L(at) \quad \text{with } L_{a,b}^\star(\lambda) \geq \frac{\lambda^2}{2(b + \lambda a)}.$$

Proof. Using the estimates (4.20) and (4.21) we prove that

$$u(L_0^\star)^{-1}(x) + v(L_1^\star)^{-1}(x) \leq 2u(x+\sqrt{x}) + v\left(\frac{x}{3} + \sqrt{2x}\right)$$
$$= a(u,v)x + \sqrt{2x\ b(u,v)}$$

with

$$a(u,v) := \left(2u + \frac{v}{3}\right) \quad \text{and} \quad b(u,v) := (\sqrt{2}\ u + v)^2.$$

Now, using Lemma 4.11, we observe that

$$ax + \sqrt{2xb} = (L_{a,b}^\star)^{-1}(x) \quad \text{with } L_{a,b}(t) = \frac{b}{2a^2} L(at). \tag{4.23}$$

Finally, we have

$$L^\star(\lambda) = (\sqrt{\lambda+1} - 1)^2 \geq \frac{(\lambda/2)^2}{(1+\lambda/2)}$$

The r.h.s. inequality can be checked easily using the fact that

$$(\sqrt{1+2\lambda} - 1)^2 = 2\left(\frac{(1+\lambda)^2 - (1+2\lambda)}{(1+\lambda) + \sqrt{1+2\lambda}}\right)$$
$$\geq \frac{\lambda^2}{(1+\lambda)} \quad (\Leftarrow\ \sqrt{1+2\lambda} \leq (1+\lambda)).$$

This implies that

$$L_{a,b}^\star(\lambda) = \frac{b}{2a^2}\ L^\star\left(\frac{2a}{b}\lambda\right) \geq \frac{\lambda^2}{2(b+\lambda a)}.$$

This ends the proof of the lemma. ∎

4.6.2 Concentration inequalities

Now, we investigate some elementary concentration inequalities for bounded and chi-square type random variables. We also apply these results to empirical processes associated with independent random variables.

Proposition 4.14. Let A be a centered random variable such that $A \leq 1$. If we set $\sigma_A = \mathbb{E}(A^2)^{1/2}$, then for any $t \geq 0$, we have

$$L_A(t) \leq \sigma_A^2\ L_1(t). \tag{4.24}$$

In addition, the probability of the following events

$$A \leq \sigma_A^2 \ (L_1^\star)^{-1}\left(\frac{x}{\sigma_A^2}\right) \leq \frac{x}{3} + \sigma_A \ \sqrt{2x}$$

is greater than $1 - e^{-x}$, for any $x \geq 0$.

Proof:

To prove (4.24) we use the fact that the decomposition

$$\mathbb{E}(e^{tA} - 1 - A) = \mathbb{E}(L_1(tA)1_{X<0}) + \mathbb{E}(L_1(tA)1_{X\in[0,1]}).$$

Since we have

$$\forall x \leq 0 \quad L_1(tx) \leq (tx)^2/2$$

and

$$\forall x \in [0,1] \quad L_1(tx) = x^2 \sum_{n\geq 2} x^{n-2}t^n/n! \leq x^2 L_1(t)$$

we conclude that

$$\begin{aligned}\mathbb{E}(e^{tA}) &\leq 1 + \frac{t^2}{2}\ \mathbb{E}(A^2 1_{A<0}) + L_1(t)\mathbb{E}(A^2 1_{A\in[0,1]}) \\ &\leq 1 + L_1(t)\sigma_A^2 \leq e^{L_1(t)\sigma_A^2}.\end{aligned}$$

Using Lemma 4.11, we readily prove that

$$(L_A^\star)^{-1}(x) \leq \sigma_A^2 \ (L_1^\star)^{-1}\left(\frac{x}{\sigma_A^2}\right) \leq \frac{x}{3} + \sigma_A \ \sqrt{2x}.$$

This ends the proof of the proposition. ∎

Proposition 4.15. For any measurable function f, with $0 < \text{osc}\ (\text{f}) \leq a$, any $N \geq 1$, and any $t \geq 0$, we have

$$L_{\sqrt{N}V(X)(f)}(t) \leq N \ \sigma^2(f/a) \ L_1(at). \tag{4.25}$$

In addition, the probability of the following events

$$\begin{aligned}V(X)(f) &\leq a^{-1}\sigma^2(f)\sqrt{N} \ (L_1^\star)^{-1}\left(\frac{xa^2}{N\sigma^2(f)}\right) \\ &\leq \frac{xa}{3\sqrt{N}} + \sqrt{2x\sigma(f)^2}\end{aligned} \tag{4.26}$$

is greater than $1 - e^{-x}$, for any $x \geq 0$.

Proof:
Replacing f by f/a, there is no loss of generality to assume that $a = 1$. Using the same arguments as those used in the proof of Proposition 4.14, we find that

$$\log \mathbb{E}(e^{t(f(X^i)-\mu^i(f))}) \leq \mu^i([f-\mu^i(f)]^2)\ L_1(t)$$

from which we conclude that

$$\begin{aligned} L_N(t) &:= \log \mathbb{E}(e^{t\sqrt{N}\ V(X)(f)}) \\ &= \sum_{i=1}^{N} \log \mathbb{E}(e^{t(f(X^i)-\mu^i(f))}) \leq \overline{L}_N(t) := N\sigma^2(f) L_1(t) \end{aligned}$$

By Lemma 4.11, we have

$$(L_N^\star)^{-1}(x) \leq (\overline{L}_N^\star)^{-1}(x) = N\sigma^2(f)\ (L_1^\star)^{-1}\left(\frac{x}{N\sigma^2(f)}\right).$$

This ends the proof of the proposition. ∎

Proposition 4.16. For any random variable B such that

$$\mathbb{E}(|B|^m)^{1/m} \leq b(2m)^2\ c \quad \text{with} \quad c < \infty$$

for any $m \geq 1$, with the finite constants $b(m)$ defined in (1.27), we have

$$L_B(t) \leq ct + L_0(ct) \tag{4.27}$$

for any $0 \leq ct < 1/2$. In addition, the probability of the following events

$$B \leq c[1 + (L_0^\star)^{-1}(x)] \leq c[1 + 2(x + \sqrt{x})]$$

is greater than $1 - e^{-x}$, for any $x \geq 0$.

Proof:
Replacing B by B/c, there is no loss of generality to assume that $c = 1$. We recall that $b(2m)^{2m} = \mathbb{E}(U^{2m})$ for every centered Gaussian random variable with $\mathbb{E}(U^2) = 1$ and

$$\forall t \in [0, 1/2) \quad \sum_{m \geq 0} \frac{t^m}{m!}\ b(2m)^{2m} = \frac{1}{\sqrt{1-2t}} = \mathbb{E}(\exp\{tU^2\}).$$

This implies that

$$\mathbb{E}(\exp\{tB\}) \leq \sum_{m\geq 0} \frac{t^m}{m!}\, b(2m)^{2m} = \frac{1}{\sqrt{1-2t}}$$

for any $0 \leq t < 1/2$. In other words, we have

$$L_{B-1}(t) := \log \mathbb{E}(\exp\{t(B-1)\} \leq L_0(t)$$

and

$$L_B(t) = t + L_{B-1}(t) \leq t + L_0(t)$$

from which we conclude that

$$\begin{aligned} L_B^\star(\lambda) &= L_{B-1}^\star(\lambda - 1) \Rightarrow (L_B^\star)^{-1}(x) \\ &= 1 + (L_{B-1}^\star)^{-1}(x) \leq 1 + (L_0^\star)^{-1}(x). \end{aligned}$$

This ends the proof of the proposition. ∎

Remark 4.6. We end with some comments on the estimate (4.5). Using the fact that $b(m) \leq b(2m)$ (see, for instance, (1.28)) we readily deduce from (4.5) that

$$\mathbb{E}(|V(X)(f)|^m)^{\frac{1}{m}} \leq\ 6\sqrt{2}\ b(2m)^2\sigma(f)$$

for any $m \geq 1$, and for any N s.t. $2\sigma^2(f)N \geq 1$. Thus, if we set

$$B = |V(X)(f)| \quad \text{and} \quad c = 6\sqrt{2}\ \sigma(f)$$

in Proposition 4.16, we prove that for any N s.t. $2\sigma^2(f)N \geq 1$, and for any $0 \leq t < 1/(12\sqrt{2}\sigma(f))$

$$L_{|V(X)(f)|}(t) \leq 6\sqrt{2}\sigma(f)t + L_0(6\sqrt{2}\ \sigma(f)t).$$

In addition, the probability of the following events

$$\begin{aligned} |V(X)(f)| &\leq 6\sqrt{2}\sigma(f)[1 + (L_0^\star)^{-1}(x)] \\ &\leq 6\sqrt{2}\sigma(f)[1 + 2(x + \sqrt{x})] \end{aligned}$$

is greater than $1 - e^{-x}$, for any $x \geq 0$.

When N is chosen so that $2\sigma^2(f)N \geq 1$, using (4.26) we improve the above inequality. Indeed, using this concentration inequality implies that for any $f \in \mathrm{Osc}(E)$, the probability of the following events

$$V(X)(f) \leq \sqrt{2}\sigma(f)\left(\frac{x}{3} + \sqrt{x}\right) \tag{4.28}$$

is greater than $1 - e^{-x}$, for any $x \geq 0$.

4.7 Perturbation Analysis

This subsection is mainly concerned with the Proof of Theorem 4.2, and Theorem 4.4. We recall that for any second-order smooth function F on $\mathbb{R}^d$, for some $d \geq 1$, $F(m(X)(f))$ stands for the random functionals defined by

$$\begin{aligned} f \;&=\; (f_i)_{1\leq i\leq d} \in \mathrm{Osc}(E)^d \\ &\mapsto F(m(X)(f)) = F(m(X)(f_1),\ldots,m(X)(f_d)) \in \mathbb{R}. \end{aligned}$$

Both results rely on the following second order decomposition, which is of independent interest.

Proposition 4.17. For any $N \geq 1$, we have the decomposition

$$\sqrt{N}[F(m(X)(f)) - F(\mu(f))] = V(X)[D_\mu(F)(f)] + \frac{1}{\sqrt{N}}R(X)(f)$$

with a first-order functional $D_\mu(F)(f)$ defined in (4.7), and a second-order term $R(X)(f)$ such that

$$\mathbb{E}(|R(X)(f)|^m)^{1/m} \leq \frac{1}{2}\; b(2m)^2\|\nabla^2 F_f\|_1$$

for any $m \geq 1$, with the parameter defined in (4.8).

Proof:
Using a Taylor first-order expansion, we have

$$\sqrt{N}[F(m(X)(f)) - F(\mu(f))] = \nabla F(\mu(f))V(X)(f)^\top + \frac{1}{\sqrt{N}}R(X)(f)$$

with the second-order remainder term

$$R(X)(f) := \int_0^1 (1-t)V(X)(f)\nabla^2 F(tm(X)(f) + (1-t)\mu(f))\ V(X)(f)^\top dt.$$

We notice that

$$\nabla F(\mu(f))\ V(X)(f)^\top = V(X)[\nabla F(\mu(f))\ f^\top]$$

and

$$\mathrm{osc}(\nabla F(\mu(f))f^\top) \leq \sum_{i=1}^d \left|\frac{\partial F}{\partial u^i}(\mu(f))\right|.$$

It is also easily checked that

$$\mathbb{E}(|R(X)(f)|^m)^{1/m} \leq \frac{1}{2}\sum_{i,j=1}^d \sup_{\nu\in\mathcal{P}(E)}\left|\frac{\partial^2 F}{\partial u^i \partial u^j}(\nu(f))\right| \mathbb{E}(|V(X)(f_i)V(X)(f_j)|^m)^{1/m}$$

and for any $1 \leq i,j \leq d$, we have

$$\mathbb{E}(|V(X)(f_i)V(X)(f_j)|^m)^{1/m} \leq \mathbb{E}(V(X)(f_j)^{2m})^{1/(2m)}\mathbb{E}(V(X)(f_j)^{2m})^{1/(2m)} \leq b(2m)^2.$$

This ends the proof of the proposition. ∎

We are now in a position to prove Theorem 4.2.

Proof of Theorem 4.2:

We set

$$N[F(m(X)(f)) - F(\mu(f))] = A + B$$

with

$$A = \sqrt{N}\ V(X)[D_\mu(F)(f)] \quad \text{and} \quad B = R(X)(f).$$

Combining Proposition 4.15 with Proposition 4.16, if we set

$$g = D_\mu(F)(f), \quad a = \|\nabla F(\mu(f))\|_1 \quad \text{and} \quad c = \frac{1}{2}\|\nabla^2 F_f\|_1$$

then we have

$$L_A(t) \leq N\sigma^2(g/a)L_1(at)$$
$$L_B(t) = ct + L_{B-c}(t) \quad \text{with } L_{B-c}(t) \leq L_0(ct).$$

On the other hand, we have

$$(L_A^\star)^{-1}(x) \leq N\ a\ \sigma^2(g/a)(L_1^\star)^{-1}\left(\frac{x}{N\sigma^2(g/a)}\right)$$

and using the fact that

$$L_B(t) = ct + L_{B-c}(t)$$

we prove that

$$\begin{aligned} L_B^\star(\lambda) = L_{B-c}^\star(\lambda - c) \Rightarrow (L_B^\star)^{-1}(x) &= c + (L_{B-c}^\star)^{-1}(x) \\ &\leq c(1 + (L_0^\star)^{-1}(x)). \end{aligned}$$

Using Bretagnolle-Rio's lemma, we find that

$$\begin{aligned} (L_{A+B}^\star)^{-1}(x) &\leq (L_A^\star)^{-1}(x) + (L_B^\star)^{-1}(x) \\ &\leq Na^{-1}\sigma^2(g)(L_1^\star)^{-1}\left(\frac{xa^2}{N\sigma^2(g)}\right) + c(1 + (L_0^\star)^{-1}(x)). \end{aligned}$$

This ends the proof of the theorem. ∎

Now, we come to the proof of Theorem 4.4.

Proof of Theorem 4.4:
We consider the empirical processes

$$f \in \mathcal{F}_i \mapsto m(X)(f) \in \mathbb{R}$$

associated with d classes of functions $\mathcal{F}_i$, $1 \leq i \leq d$, defined in subsection 4.1. We further assume that $\|f_i\| \vee \text{osc}(f_i) \leq 1$, for any $f_i \in \mathcal{F}_i$, and we set

$$\pi_\psi(\|V(X)\|_{\mathcal{F}}) := \sup_{1 \leq i \leq d} \pi_\psi(\|V(X)\|_{\mathcal{F}_i}).$$

Using Theorem 4.3, we have that

$$\pi_\psi(\|V(X)\|_{\mathcal{F}}) \leq 12^2 \int_0^2 \sqrt{\log(8 + \mathcal{N}(\mathcal{F}, \epsilon)^2)}d\epsilon$$

with

$$\mathcal{N}(\mathcal{F},\epsilon) := \sup_{1\le i\le d} \mathcal{N}(\mathcal{F}_i,\epsilon).$$

Using Proposition 4.17, for any collection of functions

$$f = (f_i)_{1\le i\le d} \in \mathcal{F} := \prod_{i=1}^{d} \mathcal{F}_i$$

we have

$$\begin{aligned}\sqrt{N}\sup_{f\in\mathcal{F}} &|F(m(X)(f)) - F(\mu(f))| \\ &\le \|\nabla F_\mu\|_\infty \sum_{i=1}^{d} \|V(X)\|_{\mathcal{F}_i} + \frac{d}{2\sqrt{N}}\|\nabla^2 F\|_\infty \sum_{i=1}^{d} \|V(X)\|_{\mathcal{F}_i}^2.\end{aligned}$$

If we set

$$A := \|\nabla F_\mu\|_\infty \sum_{i=1}^{d} \|V(X)\|_{\mathcal{F}_i}$$

then we find that

$$\pi_\psi(A) \le \|\nabla F_\mu\|_\infty \sum_{i=1}^{d} \pi_\psi(\|V(X)\|_{\mathcal{F}_i}).$$

By Lemma 4.6, this implies that

$$\mathbb{E}(e^{tA}) \le (1 + t\pi_\psi(A))\ e^{(t\pi_\psi(A))^2} \le e^{at+\frac{1}{2}t^2 b}$$

with $b = 2a^2$ and

$$a = \pi_\psi(A) \le \|\nabla F_\mu\|_\infty \sum_{i=1}^{d} \pi_\psi(\|V(X)\|_{\mathcal{F}_i}).$$

Notice that

$$L_{A-a}(t) \le L(t) = \frac{1}{2}t^2 b.$$

Recalling that

$$L^\star(\lambda) = \frac{\lambda^2}{2b} \quad \text{and} \quad (L^\star)^{-1}(x) = \sqrt{2bx}$$

we conclude that

$$\begin{aligned}(L_A^\star)^{-1}(x) &= a + (L_{A-a}^\star)^{-1}(x) \\ &\leq a + \sqrt{2bx} = \pi_\psi(A)(1 + 2\sqrt{x}).\end{aligned}$$

Now, we come to the analysis of the second-order term defined by

$$B = \frac{d}{2\sqrt{N}} \|\nabla^2 F\|_\infty \sum_{i=1}^{d} \|V(X)\|_{\mathcal{F}_i}^2.$$

Using the inequality

$$\left(\sum_{i=1}^{d} a_i \right)^m \leq d^{m-1} \sum_{i=1}^{d} a_i^m$$

which is valid for any $d \geq 1$, any $m \geq 1$, and any sequence of real numbers $(a_i)_{1 \leq i \leq d} \in \mathbb{R}_+^d$, we prove that

$$\mathbb{E}(B^m) \leq \beta^m d^{m-1} \sum_{i=1}^{d} \mathbb{E}(\|V(X)\|_{\mathcal{F}_i}^{2m})$$

with

$$\beta := \frac{d}{2\sqrt{N}} \|\nabla^2 F\|_\infty.$$

Combining Lemma 4.6 with Theorem 4.3, we conclude that

$$\mathbb{E}(B^m) \leq m! \ (\beta \ d \ \pi_\psi(\|V(X)\|_{\mathcal{F}})^2)^m.$$

If we set

$$b := \beta \ d \ \pi_\psi(\|V(X)\|_{\mathcal{F}})^2$$

then we have that

$$\mathbb{E}(e^{tB}) \leq \sum_{m \geq 0} (bt)^m = \frac{1}{1 - bt} = e^{bt} \times e^{2L_0(bt/2)}$$

for any $0 \leq t < 1/b$, with the convex increasing function L_0 introduced on page 322, so that

$$2L_0(bt/2) = -bt - \log(1 - bt).$$

Using lemma 4.11, we prove that

$$L_{B-b}(t) \leq 2L_0(bt/2)$$

and

$$\begin{aligned}(L_B^\star)^{-1}(x) &= b + (L_{B-b}^\star)^{-1}(x) \\ &\leq b\left(1 + (L_0^\star)^{-1}\left(\frac{x}{2}\right)\right) \\ &= \frac{1}{2\sqrt{N}}\|\nabla^2 F\|_\infty (d\ \pi_\psi(\|V(X)\|_{\mathcal{F}}))^2 \left(1 + (L_0^\star)^{-1}\left(\frac{x}{2}\right)\right).\end{aligned}$$

Finally, using Bretagnolle-Rio's lemma, we prove that

$$\begin{aligned}&(L_{A+B}^\star)^{-1}(x) \\ &\quad \leq d\pi_\psi(\|V(X)\|_{\mathcal{F}}) \left[\|\nabla F_\mu\|_\infty (1 + 2\sqrt{x})\right. \\ &\qquad \left. + \frac{1}{2\sqrt{N}}\|\nabla^2 F\|_\infty (d\pi_\psi(\|V(X)\|_{\mathcal{F}})) \left(1 + (L_0^\star)^{-1}\left(\frac{x}{2}\right)\right)\right].\end{aligned}$$

This ends the Proof of the Theorem 4.4. ∎

5

Interacting Empirical Processes

5.1 Introduction

Section 5 is concerned with the concentration analysis of sequences of empirical processes associated with conditionally independent random variables.

In preparation for the work in Section 6 on the collection of Feynman-Kac particle models introduced in subsection 1.4, we consider a general class of interaction particle processes with nonnecessarily mean field type dependency.

First, we analyze the concentration properties of integrals of local sampling error sequences, with general random but predictable test functions. These results will be used to analyze the concentration properties of the first-order fluctuation terms of the particle models.

We also present a stochastic perturbation technique to analyze the second-order type decompositions. We consider finite marginal models and empirical processes. We close the section with an analysis of the covering numbers and the entropy parameters of linear transformation of classes of functions.

This subsection ends with a precise description of the main mathematical objects we shall analyze further in the section.

We let $X_n^{(N)} = (X_n^{(N,i)})_{1 \leq i \leq N}$ be a Markov chain on some product state spaces E_n^N, for some $N \geq 1$. We also let $\mathcal{G}_n^N$ be the increasing σ-field generated by the random sequence $(X_p^{(N)})_{0 \leq p \leq n}$. We further assume that $(X_n^{(N,i)})_{1 \leq i \leq N}$ are conditionally independent, given $\mathcal{G}_{n-1}^N$.

As is traditional, when there is no possible confusion, we simplify notation and suppress the index $(.)^{(N)}$ so that we write $(X_n, X_n^i, \mathcal{G}_n)$ instead of $(X_n^{(N)}, X_n^{(N,i)}, \mathcal{G}_n^N)$.

In this simplified notation, we also denote by μ_n^i the conditional distribution of the random state X_n^i given the $\mathcal{G}_{n-1}$; that is, we have that

$$\mu_n^i = \mathrm{Law}(X_n^i \mid \mathcal{G}_{n-1}).$$

Notice that the conditional distributions

$$\mu_n := \frac{1}{N} \sum_{i=1}^{N} \mu_n^i$$

represent the local conditional mean of the occupation measures

$$m(X_n) := \frac{1}{N} \sum_{i=1}^{N} \delta_{X_n^i}.$$

At this level of generality, we cannot obtain any kind of concentration properties for the deviations of the occupation measures $m(X_n)$ around some deterministic limiting value.

In Section 6, dedicated to particle approximations of Feynman-Kac measures η_n, we shall deal with mean field type random measures μ_n^i, in the sense that the randomness only depends on the location of the random state X_{n-1}^i and on the current occupation measure $m(X_{n-1})$. In this situation, the fluctuation of $m(X_n)$ around the limiting deterministic measures η_n will be expressed in terms of second-order Taylor's type expansions w.r.t. the local sampling errors

$$V(X_p) = \sqrt{N}(m(X_p) - \mu_p)$$

from the origin $p = 0$, up to the current time $p = n$.

The first-order terms will be expressed in terms of integral formulae of predictable functions f_p w.r.t. the local sampling error measures $V(X_p)$. These stochastic first-order expansions are defined below.

Definition 5.1. For any sequence of $\mathcal{G}_{n-1}$-measurable random function $f_n \in \mathrm{Osc}(E_n)$, and any numbers $a_n \in \mathbb{R}_+$, we set

$$V_n(X)(f) = \sum_{p=0}^{n} a_p \; V(X_p)(f_p). \tag{5.1}$$

For any $\mathcal{G}_{n-1}$-measurable random function $f_n \in \mathrm{Osc}(E_n)$, we have

$$\mathbb{E}(V(X_n)(f_n)|\mathcal{G}_{n-1}) = 0$$

$$\mathbb{E}(V(X_n)(f_n)^2|\mathcal{G}_{n-1}) = \sigma_n^N(f_n)^2 := \frac{1}{N}\sum_{i=1}^{N} \mu_n^i([f_n - \mu_n^i(f_n)]^2).$$

We also assume that we have an almost sure estimate

$$\sup_{N\geq 1} \sigma_n^N(f_n)^2 \leq \sigma_n^2 \text{ for some positive constant } \sigma_n^2 \leq 1. \tag{5.2}$$

5.2 Finite Marginal Models

We will now derive a quantitative contraction inequality for the general random fields models of the following form

$$W_n(X)(f) = V_n(X)(f) + \frac{1}{\sqrt{N}} \; R_n(X)(f) \tag{5.3}$$

with $V_n(X)(f)$ defined in (5.1), and a second-order term such that

$$\mathbb{E}(|R_n(X)(f)|^m)^{1/m} \leq b(2m)^2 c_n$$

for any $m \geq 1$, for some finite constant $c_n < \infty$ whose values only depend on the parameter n.

For a null remainder term $R_n(X)(f) = 0$, these concentration properties are easily derived using Proposition 4.15.

Proposition 5.1. We let $V_n(X)(f)$ be the random field sequence defined in (5.1). For any $t \geq 0$, we have that

$$L_{\sqrt{N}V_n(X)(f)}(t) \leq N \; \overline{\sigma}_n^2 \; L_1(ta_n^\star)$$

with the parameters

$$\overline{\sigma}_n^2 := \sum_{0\le p\le n} \sigma_p^2 \quad \text{and} \quad a_n^\star := \max_{0\le p\le n} a_p.$$

In addition, the probability of the following events

$$\begin{aligned} V_n(X)(f) &\le \sqrt{N} a_n^\star \overline{\sigma}_n^2 (L_1^\star)^{-1}\left(\frac{x}{N\overline{\sigma}_n^2}\right) \\ &\le a_n^\star\left(\frac{x}{3\sqrt{N}} + \sqrt{2\overline{\sigma}_n^2 x}\right) \end{aligned}$$

is greater than $1-e^{-x}$, for any $x\ge 0$.

Proof:

By Proposition 4.15, we have

$$\mathbb{E}(e^{t\sqrt{N}V_n(X)(f)}|\mathcal{G}_{n-1}) = e^{t\sqrt{N}V_{n-1}(X)(f)}\ \mathbb{E}(e^{(ta_n)\sqrt{N}V(X_n)(f_n)}|\mathcal{G}_{n-1})$$

with

$$\log \mathbb{E}(e^{(ta_n)\sqrt{N}V(X_n)(f_n)}|\mathcal{G}_{n-1}) \le N\sigma_n^2\ L_1(ta_n^\star).$$

This clearly implies that

$$L_{\sqrt{N}V_n(X)(f)}(t) \le \overline{L}_1(t) := N\ \overline{\sigma}_n^2 L_1(ta_n^\star).$$

Using Lemma 4.11, we conclude that

$$\begin{aligned} (L^\star_{\sqrt{N}V_n(X)(f)})^{-1}(x) &\le (\overline{L}_1^\star)^{-1}(x) \\ &= Na_n^\star\overline{\sigma}_n^2\ (L_1^\star)^{-1}\left(\frac{x}{N\overline{\sigma}_n^2}\right). \end{aligned}$$

The last assertion is a direct consequence of (4.21). This ends the proof of the proposition. ■

Theorem 5.2. We let $W_n(X)(f)$ be the random field sequence defined in (5.3).

In this situation, the probability of the events

$$\sqrt{N}\ W_n(X)(f) \le c_n\ (1+(L_0^\star)^{-1}(x)) + Na_n^\star\ \overline{\sigma}_n^2\ (L_1^\star)^{-1}\left(\frac{x}{N\overline{\sigma}_n^2}\right)$$

is greater than $1-e^{-x}$, for any $x\ge 0$. In the above display, $\overline{\sigma}_n$ stands for the variance parameter definition in Proposition 5.1.

Proof:
We set $\sqrt{N}\ W_n(X)(f) = A_n + B_n$, with

$$A_n = \sqrt{N} V_n(X)(f) \quad \text{and} \quad B_n = R_n(X)(f).$$

By Proposition 4.16 and Proposition 5.1, we have

$$L_{A_n}(t) \leq \overline{L}_{A_n}(t) := N\ \overline{\sigma}_n^2(f)\ L_1(t a_n^\star)$$

and

$$L_{B_n - c_n}(t) \leq \overline{L}_{B_n - c_n}(t) := L_0(c_n t).$$

We recall that

$$L_{B_n}(t) = c_n t + L_{B_n - c_n}(t) \Rightarrow L^\star_{B_n}(\lambda) = L^\star_{B_n - c_n}(\lambda - c_n).$$

Using Lemma 4.11, we also have that

$$\begin{aligned}(\overline{L}^\star_{B_n})^{-1}(x) &= c_n + (L^\star_{B_n - c_n})^{-1}(x) \\ &\leq c_n + (\overline{L}^\star_{B_n - c_n})^{-1}(x) = c_n(1 + (L_0^\star)^{-1}(x)).\end{aligned}$$

In the same vein, arguing as in the end of the proof of Proposition 5.1, we have

$$(L^\star_{\sqrt{N} V_n(X)(f)})^{-1}(x) \leq N a_n^\star \overline{\sigma}_n^2 (L_1^\star)^{-1}\left(\frac{x}{N\overline{\sigma}_n^2}\right).$$

The end of the proof is now a direct consequence of the Bretagnolle-Rio's lemma. This ends the proof of the theorem. ∎

5.3 Empirical Processes

We let $V_n(X)$ be the random field sequence defined in (5.1), and we consider a sequence of classes of $\mathcal{G}_{n-1}$-measurable random functions $\mathcal{F}_n$, such that $\|f_n\| \vee \mathrm{osc}(f_n) \leq 1$, for any $f_n \in \mathcal{F}_n$.

Definition 5.2. For any $f = (f_n)_{n\geq 0} \in \mathcal{F} := (\mathcal{F}_n)_{n\geq 0}$, and any sequence of numbers $a = (a_n)_{n\geq 0} \in \mathbb{R}_+^{\mathbb{N}}$, we set

$$V_n(X)(f) = \sum_{p=0}^{n} a_p\ V(X_p)(f_p) \quad \text{and} \quad \|V_n(X)\|_{\mathcal{F}} = \sup_{f\in\mathcal{F}} |V_n(X)(f)|.$$

We further assume that for any $n \geq 0$, and any $\epsilon > 0$, we have an almost sure estimate

$$\mathcal{N}(\mathcal{F}_n, \epsilon) \leq \mathcal{N}_n(\epsilon) \tag{5.4}$$

for some non increasing function $\mathcal{N}_n(\epsilon)$ such that

$$b_n := 12^2 \int_0^2 \sqrt{\log\left(8 + \mathcal{N}_n(\epsilon)^2\right)}\, d\epsilon < \infty.$$

In this situation, we have

$$\pi_\psi(\|V_n(X)\|_{\mathcal{F}}) \leq \sum_{p=0}^{n} a_p\ \pi_\psi(\|V(X_p)\|_{\mathcal{F}_p}).$$

Using Theorem 4.3, given $\mathcal{G}_{n-1}$ we have the almost sure upper bound

$$\pi_\psi(\|V(X_p)\|_{\mathcal{F}_p}) \leq 12^2 \int_0^2 \sqrt{\log(8 + \mathcal{N}(\mathcal{F}_p, \epsilon)^2)} d\epsilon \leq .b_p$$

Combining Lemma 4.5 and Lemma 4.6, we readily prove the following theorem.

Theorem 5.3. For any classes of $\mathcal{G}_{n-1}$-measurable random functions $\mathcal{F}_n$ satisfying the entropy condition (5.4), we have

$$\pi_\psi(\|V_n(X)\|_{\mathcal{F}}) \leq c_n := \sum_{p=0}^{n} a_p b_p.$$

In particular, the probability of the events

$$\|V_n(X)\|_{\mathcal{F}} \leq c_n \sqrt{x + \log 2}$$

is greater than $1 - e^{-x}$, for any $x \geq 0$.

Next, we consider classes of nonrandom functions $\mathcal{F} = (\mathcal{F}_n)_{n \geq 0}$. We further assume that $\|f_n\| \vee \mathrm{osc}(f_n) \leq 1$, for any $f_n \in \mathcal{F}_n$, and

$$I_1(\mathcal{F}) := 12^2 \int_0^2 \sqrt{\log\left(8 + \mathcal{N}(\mathcal{F}, \epsilon)^2\right)}\, d\epsilon < \infty$$

with

$$\mathcal{N}(\mathcal{F}, \epsilon) = \sup_{n \geq 0} \mathcal{N}(\mathcal{F}_n, \epsilon) < \infty.$$

Theorem 5.4. We let $W_n(X)(f)$, $f \in \mathcal{F}$, be the random field sequence defined by

$$W_n(X)(f) = V_n(X)(f) + \frac{1}{\sqrt{N}} R_n(X)(f)$$

with a second-order term such that

$$\mathbb{E}(\sup_{f \in \mathcal{F}} |R_n(X)(f)|^m) \leq m! \; c_n^m$$

for any $m \geq 1$, for some finite constant $c_n < \infty$ whose values only depend on the parameter n. In this situation, the probability of the events

$$\|W_n(X)\|_{\mathcal{F}} \leq \left[\sum_{p=0}^{n} a_p \right] I_1(\mathcal{F})(1 + 2\sqrt{x}) + \frac{c_n}{\sqrt{N}} \left(1 + (L_0^\star)^{-1} \left(\frac{x}{2} \right) \right)$$

is greater than $1 - e^{-x}$, for any $x \geq 0$.

Proof:
We set $\sqrt{N}\|W_n(X)\|_{\mathcal{F}} \leq A_n + B_n$, with

$$A_n = \sqrt{N}\|V_n(X)\|_{\mathcal{F}} \quad \text{and} \quad B_n = \sup_{f \in \mathcal{F}} |R_n(X)(f)|.$$

Using the fact that

$$\sup_{f \in \mathcal{F}} |V_n(X)(f)| \leq \sum_{p=0}^{n} a_p \|V(X_p)\|_{\mathcal{F}_p}$$

by Lemma 4.5, we have

$$\pi_\psi(\|V_n(X)\|_{\mathcal{F}}) \leq \sum_{p=0}^{n} a_p \pi_\psi(\|V(X_p)\|_{\mathcal{F}_p}).$$

Using Theorem 4.3, we also have that

$$\pi_\psi(\|V(X_p)\|_{\mathcal{F}_p}) \leq I_1(\mathcal{F}) := 12^2 \int_0^2 \sqrt{\log(8 + \mathcal{N}(\mathcal{F}, \epsilon)^2)} d\epsilon$$

with

$$\mathcal{N}(\mathcal{F},\epsilon) = \sup_{n\geq 0}\mathcal{N}(\mathcal{F}_n,\epsilon).$$

This implies that

$$\pi_\psi(A_n) \leq \overline{a}_n\ \sqrt{N}\ I_1(\mathcal{F}) \quad \text{with } \overline{a}_n := \sum_{p=0}^{n} a_p.$$

By Lemma 4.6, we have

$$\mathbb{E}(e^{tA_n}) \leq (1 + t\pi_\psi(A_n))\ e^{(t\pi_\psi(A_n))^2} \leq e^{\alpha_n t + \frac{1}{2}t^2\beta_n}$$

with

$$\beta_n = 2\alpha_n^2 \quad \text{and} \quad \alpha_n = \pi_\psi(A_n).$$

Notice that

$$L_{A_n-\alpha_n}(t) \leq L_n(t) := \frac{1}{2}t^2\beta_n.$$

Recalling that

$$L_n^\star(\lambda) = \frac{\lambda^2}{2\beta_n} \quad \text{and} \quad (L_n^\star)^{-1}(x) = \sqrt{2\beta_n x}$$

we conclude that

$$\begin{aligned}(L_{A_n}^\star)^{-1}(x) &= \alpha_n + (L_{A_n-\alpha_n}^\star)^{-1}(x)\\ &\leq \alpha_n + \sqrt{2\beta_n x} = \pi_\psi(A_n)(1+2\sqrt{x}).\end{aligned}$$

On the other hand, under our assumption, we also have that

$$\mathbb{E}(e^{tB_n}) \leq \sum_{m\geq 0}(c_n t)^m = \frac{1}{1-c_n t} = e^{c_n t} \times e^{2L_0(c_n t/2)}$$

for any $0 \leq t < 1/c_n$ with the convex increasing function L_0 introduced on page 322, so that

$$2L_0(c_n t/2) = -c_n t - \log(1 - c_n t).$$

Using Lemma 4.11, we conclude that

$$L_{B_n-c_n}(t) \leq 2L_0(c_n t/2)$$

and

$$(L^{\star}_{B_n})^{-1}(x) = c_n + (L^{\star}_{B_n - c_n})^{-1}(x)$$
$$\leq c_n \left(1 + (L^{\star}_0)^{-1}\left(\frac{x}{2}\right)\right).$$

The end of the proof is now a direct consequence of the Bretagnolle-Rio lemma. ∎

5.4 Covering Numbers and Entropy Methods

Here, we derive some properties of covering numbers for some classes of functions. These two results are key to derive uniform concentration inequalities w.r.t. the time parameters for Feynman-Kac particle models. This subject is investigated in Section 6.

We let $(E_n, \mathcal{E}_n)_{n=0,1}$ be a pair of measurable state spaces, and $\mathcal{F}$ be a separable collection of measurable functions $f : E_1 \to \mathbb{R}$ such that $\|f\| \leq 1$ and $\mathrm{osc}(f) \leq 1$.

We consider a Markov transition $M(x_0, dx_1)$ from E_0 into E_1, a probability measure μ on E_0, and a function G from E_0 into $[0,1]$. We associate with these objects the class of functions

$$G \cdot M(\mathcal{F}) = \{G\ M(f) : f \in \mathcal{F}\}$$

and

$$G \cdot (M - \mu M)(\mathcal{F}) = \{G[M(f) - \mu M(f)] : f \in \mathcal{F}\}.$$

Lemma 5.5. For any $\epsilon > 0$, we have

$$\mathcal{N}[G \cdot M(\mathcal{F}), \epsilon] \leq \mathcal{N}(\mathcal{F}, \epsilon).$$

Proof:
For any probability measure η on E_0, we let $\{f_1, \ldots, f_{n_\epsilon}\}$ be the centers of $n_\epsilon = \mathcal{N}(\mathcal{F}, \mathbb{L}_2(\eta), \epsilon)$

There exist $\mathbb{L}_2(\eta)$-balls of radius at most ϵ covering $\mathcal{F}$. For any $f \in \mathcal{F}$, there exists some $1 \leq i \leq n_\epsilon$ such that

$$\eta([G(f - f_i)]^2)^{1/2} \leq \eta([(f - f_i)]^2)^{1/2} \leq \epsilon.$$

This implies that

$$\mathcal{N}(G \cdot \mathcal{F}, \mathbb{L}_2(\eta), \epsilon) \leq \mathcal{N}(\mathcal{F}, \mathbb{L}_2(\eta), \epsilon).$$

In much the same way, we let $\{f_1, \ldots, f_{n_\epsilon}\}$ be the $n_\epsilon = \mathcal{N}(\mathcal{F}, \mathbb{L}_2(\eta M), \epsilon)$ centers of $\mathbb{L}_2(\eta M)$-balls of radius at most ϵ covering $\mathcal{F}$. In this situation, for any $f \in \mathcal{F}$, there exists some $1 \leq i \leq n_\epsilon$ such that

$$\eta([(M(f) - M(f_i))]^2)^{1/2} \leq \eta M([(f - f_i)]^2)^{1/2} \leq \epsilon.$$

This implies that

$$\mathcal{N}(M(\mathcal{F}), \mathbb{L}_2(\eta), \epsilon) \leq \mathcal{N}(\mathcal{F}, \mathbb{L}_2(\eta M), \epsilon).$$

This ends the proof of the lemma. ∎

One of the simplest ways to control the covering numbers of the second class of functions is to assume that M satisfies the following condition $M(x, dy) \geq \delta \nu(dy)$, for any $x \in E_0$, and for some measure ν, and some $\delta \in [0, 1]$. Indeed, in this situation we observe that

$$M_\delta(x, dy) = \frac{M(x, dy) - \delta \nu(dy)}{1 - \delta}$$

is a Markov transition and

$$(1 - \delta)[M_\delta(f)(x) - M_\delta(f)(y)] = [M(f)(x) - M(f)(y)].$$

This implies that

$$(1 - \delta)[M_\delta(f)(x) - \mu M_\delta(f)] = [M(f)(x) - \mu M(f)]$$

and

$$\eta[(M(f)(x) - \mu M(f))^2] \leq 2(1 - \delta)\ \eta M_{\delta,\mu}(|f|^2)$$

with the Markov transition

$$M_{\delta,\mu}(x, dy) = \frac{1}{2}[M_\delta(x, dy) + \mu M_\delta(x, dy)].$$

We let $\{f_1, \ldots, f_{n_\epsilon}\}$ be the $n_\epsilon = \mathcal{N}(\mathcal{F}, \mathbb{L}_2(\eta M_{\delta,\mu}), \epsilon/2)$ centers of $\mathbb{L}_2(\eta M_{\delta,\mu})$-balls of radius at most ϵ covering $\mathcal{F}$. If we set

$$\overline{f} = M(f) - \mu M(f) \quad \text{and} \quad \overline{f}_i = M(f_i) - \mu M(f_i)$$

then we find that

$$\overline{f} - \overline{f}_i = M(f - f_i) - \mu M(f - f_i)$$

from which we prove that

$$\eta[(\overline{f} - \overline{f}_i)^2]^{1/2} \leq 2(1-\delta)[\eta M_{\delta,\mu}(|f - f_i|^2)]^{1/2}.$$

We conclude that

$$\mathcal{N}((M - \mu M)(\mathcal{F}), \mathbb{L}_2(\eta), 2\epsilon(1-\delta)) \leq \mathcal{N}(\mathcal{F}, \mathbb{L}_2(\eta M_{\delta,\mu}), \epsilon)$$

and therefore

$$\mathcal{N}((M - \mu M)(\mathcal{F}), 2\epsilon(1-\delta)) \leq \mathcal{N}(\mathcal{F}, \epsilon)$$

or, equivalently,

$$\mathcal{N}\left(\frac{1}{1-\delta}\ (M - \mu M)(\mathcal{F}), \epsilon\right) \leq \mathcal{N}(\mathcal{F}, \epsilon/2).$$

In more general situations, we quote the following result.

Lemma 5.6. For any $\epsilon > 0$, we have

$$\mathcal{N}[G \cdot (M - \mu M)(\mathcal{F}), \quad 2\epsilon\beta(M)] \leq \mathcal{N}(\mathcal{F}, \epsilon).$$

Proof:
We consider a Hahn-Jordan orthogonal decomposition

$$(M(x, dy) - \mu M(dy)) = M_\mu^+(x, dy) - M_\mu^-(x, dy)$$

with

$$M_\mu^+(x, dy) = (M(x, .) - \mu M)^+ \quad \text{and} \quad M_\mu^-(x, dy) = (M(x, .) - \mu M)^-$$

with

$$\|M(x, .) - \mu M\|_{\text{tv}} = M_\mu^+(x, E_1) = M_\mu^-(x, E_1) \leq \beta(M).$$

By construction, we have

$$M(f)(x) - \mu M(f) = M_\mu^+(x, E_1)\ (\overline{M}_\mu^+(f)(x) - \overline{M}_\mu^-(f)(x))$$

with

$$\overline{M}_\mu^+(x,dy) := \frac{M_\mu^+(x,dy)}{M_\mu^+(x,E_1)} \quad \text{and} \quad \overline{M}_\mu^-(x,dy) := \frac{M_\mu^-(x,dy)}{M_\mu^-(x,E_1)}.$$

This implies that

$$|M(f)(x) - \mu M(f)| \leq 2\beta(M)\ \overline{M}_\mu(|f|)(x)$$

with

$$\overline{M}_\mu(x,dy) = \frac{1}{2}(\overline{M}_\mu^+(x,dy) + \overline{M}_\mu^-(x,dy)).$$

One concludes that

$$\eta[(M(f)(x) - \mu M(f))^2]^{1/2} \leq 2\beta(M)[\eta\overline{M}_\mu(|f|^2)]^{1/2}.$$

We let $\{f_1,\ldots,f_{n_\epsilon}\}$ be the $n_\epsilon = \mathcal{N}(\mathcal{F},\mathbb{L}_2(\eta\overline{M}_\mu),\epsilon/2)$ centers of $\mathbb{L}_2(\eta\overline{M}_\mu)$-balls of radius at most ϵ covering $\mathcal{F}$.

If we set

$$\overline{f} = M(f) - \mu M(f) \quad \text{and} \quad \overline{f}_i = M(f_i) - \mu M(f_i)$$

then we find that

$$\overline{f} - \overline{f}_i = M(f - f_i) - \mu M(f - f_i)$$

from which we prove that

$$\eta[(\overline{f} - \overline{f}_i)^2]^{1/2} \leq 2\beta(M)[\eta\overline{M}_\mu(|f - f_i|^2)]^{1/2}.$$

In this situation, for any $\overline{f} \in (M - \mu M)(\mathcal{F})$, there exists some $1 \leq i \leq n_\epsilon$ such that

$$\eta[(\overline{f} - \overline{f}_i)^2]^{1/2} \leq \beta(M)\ \epsilon.$$

We conclude that

$$\mathcal{N}((M - \mu M)(\mathcal{F}),\mathbb{L}_2(\eta),\epsilon\beta(M)) \leq \mathcal{N}(\mathcal{F},\mathbb{L}_2(\eta\overline{M}_\mu),\epsilon/2)$$

and therefore

$$\begin{aligned}\mathcal{N}(G \cdot (M - \mu M)(\mathcal{F}),\epsilon\beta(M)) &\leq \mathcal{N}((M - \mu M)(\mathcal{F}),\epsilon\beta(M)) \\ &\leq \mathcal{N}(\mathcal{F},\epsilon/2).\end{aligned}$$

This ends the proof of the lemma. ∎

6

Feynman-Kac Particle Processes

6.1 Introduction

In this section, we investigate the concentration properties of the collection of Feynman-Kac particle measures introduced in subsection 1.4, in terms of the contraction parameters $\tau_{k,l}(n)$ and $\overline{\tau}_{k,l}(m)$ introduced in Definition 3.3, and in Corollary 3.5.

First, we present some basic first-order decompositions of the Boltzmann-Gibbs transformation associated with some regular potential function.

Then, we combine the semigroup techniques developed in section 3, with a stochastic perturbation analysis to derive first-order integral expansions in terms of local random fields and Feynman-Kac transport operators.

In subsection 6.4, we combine these key formulae with the concentration analysis of interacting empirical processes developed in Section 5. We derive quantitative concentration estimates for finite marginal models, as well as for empirical processes w.r.t. some classes of functions. Subsections 6.5 and 6.6 are devoted to particle free energy models, and backward particle Markov models, respectively.

6.2 First-Order Expansions

For any positive potential function G, any measures μ and ν, and any function f on E, we have

$$\begin{aligned}[\Psi_G(\mu) - \Psi_G(\nu)](f) &= \frac{1}{\mu(G_\nu)}(\mu-\nu)(d_\nu\Psi_G(f)) \\ &= \left(1 - \frac{1}{\mu(G_\nu)}\ (\mu-\nu)(G_\nu)\right)(\mu-\nu)(d_\nu\Psi_G(f)) \end{aligned} \tag{6.1}$$

with the functions

$$d_\nu\Psi_G(f) := G_\nu\ (f - \Psi_G(\nu)(f)) \quad \text{and} \quad G_\nu := G/\nu(G). \tag{6.2}$$

Notice that

$$|[\Psi_G(\mu) - \Psi_G(\nu)](f)| \leq g|(\mu-\nu)(d_\nu\Psi_G(f))|$$

and

$$\|d_\nu\Psi_G(f)\| \leq g\ \mathrm{osc}(f) \quad \text{with } g := \sup_{x,y}(G(x)/G(y)).$$

It is also important to observe that

$$\begin{aligned}|[\Psi_G(\mu) - \Psi_G(\nu)](f)| &\leq \frac{1}{\mu(G')}|(\mu-\nu)(d'_\nu\Psi_G(f))| \\ &\leq g|(\mu-\nu)(d'_\nu\Psi_G(f))|\end{aligned}$$

with the integral operator $d'_\nu\Psi_G$ from Osc(E) into itself defined by

$$d'_\nu\Psi_G(f) := G'(f - \Psi_G(\nu)(f)) \quad \text{and} \quad G' := G/\|G\|.$$

Using Lemma 5.6, we readily prove the following lemma.

Lemma 6.1. We let $\mathcal{F}$ be a separable collection of measurable functions $f : E' \to \mathbb{R}$ on some possibly different state space E', and such that $\|f\| \leq 1$ and $\mathrm{osc}(f) \leq 1$. For any Markov transition M from E into E', we set

$$d'_\nu\Psi_G M(\mathcal{F}) := \{d'_\nu\Psi_G(M(f)) : f \in \mathcal{F}\}.$$

In this situation, we have the uniform estimate

$$\sup_{\nu\in\mathcal{P}(E)} \mathcal{N}[d'_\nu\Psi_G M(\mathcal{F}), 2\epsilon\beta(M)] \leq \mathcal{N}(\mathcal{F},\epsilon). \tag{6.3}$$

6.3 A Stochastic Perturbation Analysis

Mean field particle models can be thought of as a stochastic perturbation technique for solving nonlinear measure valued equations of the form

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

The random perturbation term is encapsulated into the sequence of local random sampling errors $(V_n^N)_{n\geq 0}$ given by the local perturbation equations

$$\eta_n^N = \Phi_n(\eta_{n-1}^N) + \frac{1}{\sqrt{N}}\ V_n^N.$$

One natural way to control the fluctuations and the concentration properties of the particle measures (η_n^N, γ_n^N) around their limiting values (η_n, γ_n) is to express the random fields $(W_n^{\gamma,N}, W_n^{\eta,N})$ defined by

$$\gamma_n^N = \gamma_n + \frac{1}{\sqrt{N}}\ W_n^{\gamma,N} \quad \eta_n^N = \eta_n + \frac{1}{\sqrt{N}}\ W_n^{\eta,N}$$

in terms of the empirical random fields $(V_n^N)_{n\geq 0}$.

As shown in (4.2), it is important to recall that the local sampling random fields models V_n^N belong to the class of empirical processes we analyzed in subsection 4.1. The stochastic analysis developed in Section 4 applies directly to these models. For instance, using Theorem 4.1 we have the quantitative almost sure estimate of the amplitude of the stochastic perturbations

$$\mathbb{E}(|V_n^N(f)|^m | \mathcal{G}_{n-1}^N)^{1/m} \leq b(m) \tag{6.4}$$

for any $m \geq 1$ and any test function $f \in \mathrm{Osc}(E_n)$.

The first-order expansions presented in the further development of this section will be expressed in terms of the random functions $d_{p,n}^N(f)$ and $G_{p,n}^N$, and the first-order functions $d_{p,n}(f)$ defined below.

Definition 6.1. For any $0 \leq p \leq n$, and any function f on E_n, we denote by $d_{p,n}(f)$ the function on E_p defined by

$$d_{p,n}(f) = d_{\eta_p} \Psi_{G_{p,n}}(P_{p,n}(f)).$$

For any $N \geq 1$, and any $0 \leq p \leq n$, we also denote by $G_{p,n}^N$, $d_{p,n}^N(f)$, and $d_{p,n}'^N(f)$ the $\mathcal{G}_{p-1}^N$-measurable random functions on E_p given by

$$G_{p,n}^N := \frac{G_{p,n}}{\Phi_p(\eta_{p-1}^N)(G_{p,n})} \quad d_{p,n}^N(f) := d_{\Phi_p(\eta_{p-1}^N)}\Psi_{G_{p,n}}(P_{p,n}(f))$$

and

$$d_{p,n}'^N(f) := d'_{\Phi_p(\eta_{p-1}^N)}\Psi_{G_{p,n}}(P_{p,n}(f)).$$

Notice that

$$\|G_{p,n}^N\| \leq g_{p,n} \quad \text{and} \quad \|d_{p,n}(f)\| \vee \|d_{p,n}^N(f)\| \leq g_{p,n}\ \beta(P_{p,n})$$

as well as

$$\|d_{p,n}'^N(f)\| \leq \beta(P_{p,n}) \quad \text{and} \quad \operatorname{osc}(d_{p,n}'^N(f)) \leq 2\beta(P_{p,n}).$$

As promised, the next theorem presents some key first-order decompositions, which are the progenitors for our other results. Further details on these expansions and their use in the bias and the fluctuation analysis of Feynman-Kac particle models can be found in [15, 17, 38] and [20].

Theorem 6.2. For any $0 \leq p \leq n$, and any function f on E_n, we have the decomposition

$$W_n^{\eta,N}(f) = \sum_{p=0}^{n} \frac{1}{\eta_p^N(G_{p,n}^N)}\ V_p^N(d_{p,n}^N(f)) \tag{6.5}$$

and the $\mathbb{L}_m$-mean error estimates

$$\mathbb{E}(|W_n^{\eta,N}(f)|^m)^{1/m} \leq 2\ b(m)\tau_{1,1}(n) \tag{6.6}$$

with the parameter $\tau_{1,1}(n)$ defined in (3.6).

In addition, we have

$$W_n^{\eta,N}(f) = \sum_{p=0}^{n} V_p^N[d_{p,n}(f)] + \frac{1}{\sqrt{N}} R_n^N(f) \tag{6.7}$$

with a second-order remainder term

$$R_n^N(f) := -\sum_{p=0}^{n-1} \frac{1}{\eta_p^N(\overline{G}_p)}\ W_p^{\eta,N}(\overline{G}_p)\ W_p^{\eta,N}[d_{p,n}(f)]$$

such that

$$\sup_{f\in \mathrm{Osc}(E_n)} \mathbb{E}[|R_n^N(f)|^m]^{1/m} \leq 4b(2m)^2\tau_{2,1}(n) \tag{6.8}$$

with the parameter $\tau_{2,1}(n)$ defined in (3.6).

Proof. The proof of (6.5) is based on the telescoping sum decomposition

$$\eta_n^N - \eta_n = \sum_{p=0}^{n}[\Phi_{p,n}(\eta_p^N) - \Phi_{p,n}(\Phi_p(\eta_{p-1}^N))]$$

Recalling that

$$\Phi_{p,n}(\mu) = \Psi_{G_{p,n}}(\mu)P_{p,n}$$

we prove that

$$\Phi_{p,n}(\eta_p^N) - \Phi_{p,n}(\Phi_p(\eta_{p-1}^N)) = [\Psi_{G_{p,n}}(\eta_p^N) - \Psi_{G_{p,n}}(\Phi_p(\eta_{p-1}^N))]P_{p,n}.$$

Using (6.1), we have

$$\begin{aligned}
&\sqrt{N}[\Psi_{G_{p,n}}(\eta_p^N) - \Psi_{G_{p,n}}(\Phi_p(\eta_{p-1}^N))](f)\\
&\quad = V_p^N[d_{\Phi_p(\eta_{p-1}^N)}\Psi_{G_{p,n}}(f)]\\
&\qquad - \frac{1}{\sqrt{N}}\frac{1}{\eta_p^N(G_{p,n}^N)}V_p^N(G_{p,n}^N)V_p^N[d_{\Phi_p(\eta_{p-1}^N)}\Psi_{G_{p,n}}(f)].
\end{aligned}$$

The proof of (6.7) is based on the telescoping sum decomposition

$$\eta_n^N - \eta_n = \sum_{p=0}^{n}[\eta_p^N\overline{Q}_{p,n} - \eta_{p-1}^N\overline{Q}_{p-1,n}]$$

with the convention $\eta_{-1}^N\overline{Q}_{-1,n} = \eta_0\overline{Q}_{0,n} = \eta_n$, for $p = 0$. Using the fact that

$$\eta_{p-1}^N\overline{Q}_{p-1,n}(f) = \eta_{p-1}^N(\overline{G}_{p-1}) \times \Phi_p(\eta_{p-1}^N)\overline{Q}_{p,n}(f)$$

we prove that

$$[\eta_n^N - \eta_n](f) = \sum_{p=0}^{n} [\eta_p^N - \Phi_p(\eta_{p-1}^N)]\overline{Q}_{p,n}(f) + R_n^N(f)$$

with the second-order remainder term

$$R_n^N(f) := \sum_{p=1}^{n} (1 - \eta_{p-1}^N(\overline{G}_{p-1})) \times \Phi_p(\eta_{p-1}^N)\overline{Q}_{p,n}(f).$$

Replacing f by the centered function $(f - \eta_n(f))$, and using the fact that

$$1 = \eta_{p-1}(\overline{G}_{p-1}) \quad \text{and} \quad \eta_p[d_{p,n}(f - \eta_n(f))] = 0$$

we conclude that

$$[\eta_n^N - \eta_n](f) = \sum_{p=0}^{n} [\eta_p^N - \Phi_p(\eta_{p-1}^N)](d_{p,n}(f)) + \overline{R}_n^N(f)$$

with the second-order remainder term

$$\begin{aligned}
\overline{R}_n^N(f) := & \sum_{p=1}^{n} [\eta_{p-1} - \eta_{p-1}^N](\overline{G}_{p-1}) \\
& \times [\Psi_{G_{p-1}}(\eta_{p-1}^N) - \Psi_{G_{p-1}}(\eta_{p-1})](M_p(d_{p,n}(f))) \\
= & -\frac{1}{N} \sum_{p=1}^{n} W_{p-1}^{\eta,N}(\overline{G}_{p-1}) \\
& \times \frac{1}{\eta_{p-1}^N(\overline{G}_{p-1})} W_{p-1}^{\eta,N}(d_{\eta_{p-1}} \Psi_{G_{p-1}}(M_p(d_{p,n}(f)))).
\end{aligned}$$

Finally, we observe that

$$\begin{aligned}
d_{\eta_{p-1}} \Psi_{G_{p-1}}(M_p(d_{p,n}(f))) &= \frac{G_{p-1}}{\eta_{p-1}(G_{p-1})}(M_p(d_{p,n}(f)) - \eta_p(d_{p,n}(f))) \\
&= \frac{G_{p-1}}{\eta_{p-1}(G_{p-1})} M_p(d_{p,n}(f)) \\
&= \overline{Q}_{p-1,p}(d_{p,n}(f)) = \overline{Q}_{p-1,n}(f - \eta_n(f)) \\
&= d_{p-1,n}(f).
\end{aligned}$$

This ends the proof of (6.7).

Combining (6.5) with the almost sure estimates

$$\mathbb{E}(|V_p^N(d_{p,n}^N(f_n))|^m|\mathcal{G}_{p-1}^N)^{1/m} \leq 2b(m)\|G_{p,n}^N\|\ \beta(P_{p,n})$$

for any $m \geq 1$, we easily prove (6.6). Using the fact that

$$\frac{1}{\inf_x G_{p,n}^N(x)}\mathbb{E}[|V_p^N(G_{p,n}^N)|^m|\mathcal{G}_{p-1}^N]^{1/m} \leq 2b(m)\ g_{p,n}$$

and

$$\mathbb{E}[|V_p^N[d_{p,n}^N(f)]|^m|\mathcal{G}_{p-1}^N]^{1/m} \leq 2b(m)g_{p,n}\beta(P_{p,n})$$

we prove (6.8). This ends the proof of the theorem. ∎

6.4 Concentration Inequalities

6.4.1 Finite Marginal Models

In subsection 3.6, dedicated to the variance analysis of the local sampling models, we have seen that the empirical random fields V_n^N satisfy the regularity conditions of the general interacting empirical process models $V(X_n)$ presented in subsection 5.1.

To be more precise, we have the formulae

$$\sum_{p=0}^{n} V_p^N[d_{p,n}(f)] = \sum_{p=0}^{n} a_p\ V_p^N[\delta_{p,n}(f)]$$

with the functions

$$\delta_{p,n}(f) = d_{p,n}(f)/a_p \in \mathrm{Osc}(E_p) \cap \mathcal{B}_1(E_p)$$

for any finite constants

$$a_p \geq 2 \sup_{0\leq p\leq n}(g_{p,n}\beta(P_{p,n})).$$

Therefore, if we fix the final time horizon n, with some slight abuse of notation we have the formulae

$$\sum_{p=0}^{n} V_p^N[d_{p,n}(f)] = \sum_{p=0}^{n}\ \alpha_p\ V(X_p)[f_p]$$

with

$$X_p = \xi_p = (\xi_p^i)_{1\leq i\leq N} \quad \alpha_p = \sup_{0\leq q\leq n} a_q = a_n^\star \quad \text{and} \quad f_p = \delta_{p,n}(f)/a_n^\star.$$

We also notice that

$$\begin{aligned}\mathbb{E}(V_p^N[\delta_{p,n}(f)]^2|\mathcal{G}_{p-1}^N) &\leq \frac{1}{(a_n^\star)^2}\sigma_p^2 \mathrm{osc}(d_{p,n}(f))^2 \\ &\leq \frac{4}{(a_n^\star)^2}\sigma_p^2 \|d_{p,n}(f)\|^2\end{aligned}$$

and therefore

$$\mathbb{E}(V_p^N[\delta_{p,n}(f)]^2|\mathcal{G}_{p-1}^N) \leq \frac{4}{(a_n^\star)^2}\sigma_p^2 g_{p,n}^2 \beta(P_{p,n})^2$$

with the uniform local variance parameters σ_p^2 defined in (3.20).

This shows that the regularity condition stated in (5.2) is met by replacing the parameters σ_p in the variance formula (5.2) by the constants $2\sigma_p g_{p,n}\beta(P_{p,n})/a_n^\star$, with the uniform local variance parameters σ_p defined in (3.20).

Using Theorem 5.2, we easily prove the following exponential concentration property.

Theorem 6.3. ([40]) For any $n \geq 0$, any $f \in \mathrm{Osc}(E_n)$, and any $N \geq 1$, the probability of the event

$$[\eta_n^N - \eta_n](f) \leq \frac{4\tau_{2,1}(n)}{N}(1 + (L_0^\star)^{-1}(x)) + 2b_n\overline{\sigma}_n^2(L_1^\star)^{-1}\left(\frac{x}{N\overline{\sigma}_n^2}\right)$$

is greater than $1 - e^{-x}$, for any $x \geq 0$, with

$$\overline{\sigma}_n^2 := \frac{1}{b_n^2}\sum_{0\leq p\leq n} g_{p,n}^2\beta(P_{p,n})^2\sigma_p^2$$

for any choice of $b_n \geq \kappa(n)$. In the above display, $\tau_{2,1}(n)$ and $\kappa(n)$ stands for the parameters defined in (3.6), and σ_n is the uniform local variance parameter defined in (3.20).

We illustrate the impact of this theorem with two applications. The first one is concerned with regular and stable Feynman-Kac models satisfying the regularity conditions presented in subsection 3.4. The second is concerned with the concentration properties of the genealogical tree based models developed in subsection 1.17.

In the first situation, combining corollary 3.5, with the estimates (4.20) and (4.21) we prove the following uniform concentration inequalities w.r.t. the time horizon.

Corollary 6.4. We assume that one of the regularity conditions $\mathbf{H_m(G,M)}$ stated in subsection 3.4.1 is met for some $m \geq 0$, and we set

$$p_m(x) = 4\overline{\tau}_{2,1}(m)(1 + 2(x + \sqrt{x})) + \frac{2}{3}\overline{\kappa}(m)x$$

and

$$q_m(x) = \sqrt{8\sigma^2\overline{\tau}_{2,2}(m)\ x} \quad \text{with } \sigma^2 = \sup_{n\geq 0}\sigma_n^2.$$

In the above displayed formula, $\overline{\tau}_{2,2}(m)$ and $\overline{\kappa}(m)$ stands for the parameters defined in Corollary 3.5, and σ_n is the uniform local variance parameter defined in (3.20).

In this situation, for any $n \geq 0$, any $f \in \mathrm{Osc}(E_n)$, any $N \geq 1$, and for any $x \geq 0$, the probability of the event

$$[\eta_n^N - \eta_n](f) \leq \frac{1}{N}p_m(x) + \frac{1}{\sqrt{N}}q_m(x)$$

is greater than $1 - e^{-x}$.

In the same vein, using the estimates (4.20) and (4.21), concentration inequalities for genealogical tree models can be derived easily using the estimates (3.14).

Corollary 6.5. We let η_n^N be the occupation measure of the genealogical tree model presented in (3.18). We also set $\sigma^2 = \sup_{n\geq 0}\sigma_n^2$, the supremum of which is the uniform local variance parameters σ_n^2 defined in (3.20), and

$$p_{n,m}(x) = 4(\chi_m g^m)^2(1 + 2(x + \sqrt{x})) + \frac{2}{3}\frac{\chi_m g^m}{(n+1)}x$$

and

$$q_m(x) = (\chi_m g^m)\sqrt{8\sigma^2 x}.$$

In this situation, for any $n \geq 0$, any $\mathbf{f_n} \in \mathrm{Osc}(\mathbf{E_n})$, and any $N \geq 1$, the probability of the event

$$[\eta_n^N - \mathbb{Q}_n](\mathbf{f_n}) \leq \frac{n+1}{N} p_{n,m}(x) + \sqrt{\frac{n+1}{N}} q_m(x)$$

is greater than $1 - e^{-x}$, for any $x \geq 0$.

6.4.2 Empirical Processes

The main aim of this subsection is to derive concentration inequalities for particle empirical processes. Several consequences of this general theorem are also discussed, including uniform estimates w.r.t. the time parameter, and concentration properties of genealogical particle processes.

Theorem 6.6. We let $\mathcal{F}_n$ be a separable collection of measurable functions f_n on E_n, such that $\|f_n\| \leq 1$, $\mathrm{osc}(f_n) \leq 1$, with finite entropy $I(\mathcal{F}_n) < \infty$.

$$\pi_\psi(\|W_n^{\eta,N}\|_{\mathcal{F}_n}) \leq c_{\mathcal{F}_n} \tau_{1,1}(n)$$

with the parameter $\tau_{1,1}(n)$ defined in (3.6) and

$$c_{\mathcal{F}_n} \leq 24^2 \int_0^1 \sqrt{\log\left(8 + \mathcal{N}(\mathcal{F}_n, \epsilon)^2\right)} d\epsilon. \tag{6.9}$$

In particular, for any $n \geq 0$, and any $N \geq 1$, the probability of the following event

$$\sup_{f \in \mathcal{F}_n} |\eta_n^N(f) - \eta_n(f)| \leq \frac{c_{\mathcal{F}_n}}{\sqrt{N}}\ \tau_{1,1}(n)\ \sqrt{x + \log 2}$$

is greater than $1 - e^{-x}$, for any $x \geq 0$.

Proof:
Using (6.5), for any function $f_n \in \mathrm{Osc}(E_n)$ we have the estimate

$$|W_n^{\eta,N}(f_n)| \leq 2 \sum_{p=0}^{n} g_{p,n} \beta(P_{p,n}) |V_p^N(\delta_{p,n}^N(f_n))|$$

with the $\mathcal{G}_{p-1}^N$-measurable random functions $\delta_{p,n}^N(f_n)$ on E_p defined by

$$\delta_{p,n}^N(f_n) = \frac{1}{2\beta(P_{p,n})}\ d_{p,n}'^N(f_n).$$

By construction, we have

$$\|\delta_{p,n}^N(f_n)\| \leq 1/2 \quad \text{and} \quad \delta_{p,n}^N(f_n) \in \mathrm{Osc}(E_p).$$

Using the uniform estimate (6.3), if we set

$$\mathcal{G}_{p,n}^N := \delta_{p,n}^N(\mathcal{F}_n) = \{\delta_{p,n}^N(f) : f \in \mathcal{F}_n\}$$

then we also prove the almost sure upper bound

$$\sup_{N\geq 1} \mathcal{N}[\mathcal{G}_{p,n}^N, \epsilon] \leq \mathcal{N}(\mathcal{F}_n, \epsilon/2).$$

The end of the proof is now a direct consequence of Theorem 5.3. This ends the proof of the theorem. ∎

Corollary 6.7. We consider time homogeneous Feynman-Kac models on some common measurable state space $E_n = E$. We also let $\mathcal{F}$ be a separable collection of measurable functions f on E, such that $\|f\| \leq 1$, $\mathrm{osc}(f) \leq 1$, with finite entropy $I(\mathcal{F}) < \infty$.

We also assume that one of the regularity conditions $\mathbf{H_m}(\mathbf{G},\mathbf{M})$ stated in subsection 3.4.1 is met for some $m \geq 0$. In this situation, for any $n \geq 0$, and any $N \geq 1$, the probability of the following event

$$\sup_{f\in\mathcal{F}_n} |\eta_n^N(f) - \eta_n(f)| \leq \frac{c_{\mathcal{F}}}{\sqrt{N}} \overline{\tau}_{1,1}(m)\sqrt{x + \log 2}$$

is greater than $1 - e^{-x}$, for any $x \geq 0$.

In the same vein, using the estimates (3.14), we easily prove the following corollary.

Corollary 6.8. We also assume that one of the regularity conditions $\mathbf{H_m}(\mathbf{G},\mathbf{M})$ stated in subsection 3.4.1 is met for some $m \geq 0$.

We let $\mathcal{F}_n$ be a separable collection of measurable functions $\mathbf{f_n}$ on the path space $\mathbf{E_n}$, such that $\|\mathbf{f_n}\| \leq 1$, $\mathrm{osc}(\mathbf{f_n}) \leq 1$, with finite entropy $I(\mathcal{F}_n) < \infty$.

We also let η_n^N be the occupation measure of the genealogical tree model presented in (3.18). In this situation, for any $n \geq 0$, and any $N \geq 1$, the probability of the following event

$$\sup_{\mathbf{f_n} \in \mathcal{F}_n} |\eta_n^N(\mathbf{f_n}) - \mathbb{Q}_n(\mathbf{f_n})| \leq \frac{c_{\mathcal{F}_n}}{\sqrt{N}}(n+1)\chi_m g^m \sqrt{x + \log 2}$$

is greater than $1 - e^{-x}$, for any $x \geq 0$ with the constant $c_{\mathcal{F}_n}$ defined in (6.9).

The following corollaries are a direct consequence of (4.10).

Corollary 6.9. We assume that the conditions stated in Corollary 6.7 are satisfied. When $\mathcal{F}$ stands for the indicator functions (4.9) of cells in $E = \mathbb{R}^d$, for some $d \geq 1$, the probability of the following event

$$\sup_{f \in \mathcal{F}} |\eta_n^N(f) - \eta_n(f)| \leq c\overline{\tau}_{1,1}(m)\sqrt{\frac{d}{N}(x+1)}$$

is greater than $1 - e^{-x}$, for any $x \geq 0$, for some universal constant $c < \infty$ that does not depend on the dimension.

Corollary 6.10. We assume that the conditions stated in corollary 6.8 are satisfied. When $\mathcal{F}_n$ stands for the product functions of cell indicators (4.9) in the path space $\mathbf{E_n} = (\mathbb{R}^{d_0} \times \cdots, \times \mathbb{R}^{d_n})$, for some $d_p \geq 1$, $p \geq 0$, the probability of the following event

$$\sup_{\mathbf{f_n} \in \mathcal{F}_n} |\eta_n^N(\mathbf{f_n}) - \mathbb{Q}_n(\mathbf{f_n})| \leq c(n+1)\chi_m g^m \sqrt{\frac{\sum_{0 \leq p \leq n} d_p}{N}(x+1)}$$

is greater than $1 - e^{-x}$, for any $x \geq 0$, for some universal constant $c < \infty$ that does not depend on the dimension.

6.5 Particle Free Energy Models

6.5.1 Introduction

The main aim of this subsection is to analyze the concentration properties of the particle free energy models introduced in subsection 1.4.2.

More formally, the unnormalized particle random field models discussed in this section are defined below.

Definition 6.2. We denote by $\overline{\gamma}_n^N$ the normalized models defined by the following formulae

$$\overline{\gamma}_n^N(f) = \gamma_n^N(f)/\gamma_n(\mathbb{1}) = \eta_n^N(f) \prod_{0\leq p<n} \eta_p^N(\overline{G}_p)$$

with the normalized potential functions

$$\overline{G}_n := G_n/\eta_n(G_n).$$

We also let $W_n^{\gamma,N}$ and $\overline{W}_n^{\gamma,N}$ be the random field particle models defined by

$$W_n^{\gamma,N} = \sqrt{N}[\gamma_n^N - \gamma_n] \quad \text{and} \quad \overline{W}_n^{\gamma,N} := W_n^{\gamma,N}(f)/\gamma_n(\mathbb{1}).$$

These unnormalized particle models γ_n^N have a particularly simple form. They are defined in terms of the product of empirical mean values $\eta_p^N(G_p)$ of the potential functions G_p w.r.t. the flow of normalized particle measures η_p^N after the p-th mutation stages, with $p < n$.

Thus, the concentration properties of γ_n^N should be related in some way to those of the interacting processes η_n^N developed in subsection 6.4.

To begin with, we mention that

$$\overline{\gamma}_n^N(\mathbb{1}) := \gamma_n^N(\mathbb{1})/\gamma_n(\mathbb{1}) = \prod_{0\leq p<n} \eta_p^N(\overline{G}_p) = 1 + \frac{1}{\sqrt{N}} \overline{W}_n^{\gamma,N}(\mathbb{1}).$$

For more general functions we also observe that for any function f on E_n, s.t. $\eta_n(f) = 1$, we have the decompositions

$$\begin{aligned}
&\overline{W}_n^{\gamma,N}(f) \\
&\quad = \sqrt{N}\left[\left(1 + \frac{1}{\sqrt{N}} \overline{W}_n^{\gamma,N}(\mathbb{1})\right)\left(\eta_n(f) + \frac{1}{\sqrt{N}} W_n^{\eta,N}(f)\right) - \eta_n(f)\right] \\
&\quad = \sqrt{N}\left[\left(1 + \frac{1}{\sqrt{N}}\ \overline{W}_n^{\gamma,N}(\mathbb{1})\right)\left(1 + \frac{1}{\sqrt{N}}\ W_n^{\eta,N}(f)\right) - 1\right].
\end{aligned}$$

We readily deduce the following second order decompositions of the fluctuation errors

$$\overline{W}_n^{\gamma,N}(f) = [\overline{W}_n^{\gamma,N}(\mathbb{1}) + W_n^{\eta,N}(f)] + \frac{1}{\sqrt{N}}(\overline{W}_n^{\gamma,N}(\mathbb{1})W_n^{\eta,N}(f)).$$

This decomposition allows us to reduce the concentration properties of $\overline{W}_n^{\gamma,N}(f)$ to those of $W_n^{\eta,N}(f)$ and $\overline{W}_n^{\gamma,N}(\mathbb{1})$.

Next, we provide some key decompositions of $W_n^{\gamma,N}$ in terms of the local sampling errors V_n^N, as well as a pivotal exponential formula connecting the fluctuations of the particle free energies in terms of the fluctuations of the potential empirical mean values.

Then, we derive first-order expansions, and logarithmic concentration inequalities for particle free energy ratios $\overline{\gamma}_n^N(\mathbb{1}) = \gamma_n^N(\mathbb{1})/\gamma_n(\mathbb{1})$.

6.5.2 Some Key Decomposition Formulae

This subsection is mainly concerned with the proof of the following decomposition theorem.

Theorem 6.11. For any $0 \leq p \leq n$, and any function f on E_n, we have the decompositions

$$W_n^{\gamma,N}(f) = \sum_{p=0}^{n} \gamma_p^N(\mathbb{1})\ V_p^N(Q_{p,n}(f)) \tag{6.10}$$

$$\overline{W}_n^{\gamma,N}(f) = \sum_{p=0}^{n} \overline{\gamma}_p^N(\mathbb{1})\ V_p^N(\overline{Q}_{p,n}(f)) \tag{6.11}$$

with the normalized Feynman-Kac semigroup

$$\overline{Q}_{p,n}(f) = Q_{p,n}(f)/\eta_p Q_{p,n}(\mathbb{1}).$$

In addition, we have the exponential formulae

$$\overline{W}_n^{\gamma,N}(\mathbb{1}) = \sqrt{N}\left(\exp\left\{\frac{1}{\sqrt{N}}\int_0^1 \sum_{0\leq p<n} \frac{W_p^{\eta,N}(\overline{G}_p)}{1+\frac{t}{\sqrt{N}}W_p^{\eta,N}(\overline{G}_p)}dt\right\} - 1\right). \tag{6.12}$$

Proof:
We use the telescoping sum decomposition

$$\gamma_n^N - \gamma_n = \sum_{p=0}^{n} (\gamma_p^N Q_{p,n} - \gamma_{p-1}^N Q_{p-1,n})$$

with the conventions $Q_{n,n} = Id$, for $p = n$; and $\gamma_{-1}^N Q_{-1,n} = \gamma_0 Q_{0,n}$, for $p = 0$. Using the fact that

$$\gamma_p^N(\mathbb{1}) = \gamma_{p-1}^N(G_{p-1}) \quad \text{and} \quad \gamma_{p-1}^N Q_{p-1,n}(f) = \gamma_{p-1}^N(G_{p-1} M_p(Q_{p,n}(f)))$$

we prove that

$$\gamma_{p-1}^N Q_{p-1,n} = \gamma_p^N(\mathbb{1})\ \Phi_p(\eta_{p-1}^N) Q_{p,n}.$$

The end of the proof of the first decomposition is now easily completed. We prove (6.11) using the following formulae

$$\begin{aligned} \overline{Q}_{p,n}(f)(x) &= \frac{\gamma_p(\mathbb{1})}{\gamma_n(\mathbb{1})} Q_{p,n}(f)(x) \\ &= Q_{p,n}(f)(x) \prod_{p \le q < n} \eta_q(G_q)^{-1} = \frac{Q_{p,n}(f)(x)}{\eta_p Q_{p,n}(\mathbb{1})}. \end{aligned}$$

The proof of (6.12) is based on the fact that

$$\log y - \log x = \int_0^1 \frac{(y-x)}{x + t(y-x)} dt$$

for any positive numbers x, y. Indeed, we have the formula

$$\begin{aligned} \log\left(\gamma_n^N(\mathbb{1})/\gamma_n(\mathbb{1})\right) &= \log\left(1 + \frac{1}{\sqrt{N}} \frac{W_n^{\gamma,N}(\mathbb{1})}{\gamma_n(\mathbb{1})}\right) \\ &= \sum_{0 \le p < n} \left(\log \eta_p^N(G_p) - \log \eta_p(G_p)\right) \\ &= \frac{1}{\sqrt{N}} \sum_{0 \le p < n} \int_0^1 \frac{W_p^{\eta,N}(G_p)}{\eta_p(G_p) + \frac{t}{\sqrt{N}} W_p^{\eta,N}(G_p)} dt. \end{aligned}$$

This ends the proof of the theorem. ∎

6.5.3 Concentration Inequalities

Combining the exponential formulae (6.12) with the expansions (6.11), we derive first-order decompositions for the random sequence $\sqrt{N}\ \log \overline{\gamma}_n^N(\mathbb{1})$. These expansions will be expressed in terms of the random predictable functions defined below.

Definition 6.3. We let $h_{q,n}^N$ be the random $\mathcal{G}_{q-1}^N$-measurable functions given by

$$h_{q,n}^N := \sum_{q \leq p < n} d_{q,p}^N(\overline{G}_p)$$

with the functions $d_{q,p}^N(\overline{G}_p)$ given in Definition 6.1.

Lemma 6.12. For any $n \geq 0$ and any $N \geq 1$, we have

$$\sqrt{N} \log \overline{\gamma}_n^N(\mathbb{1}) = \sum_{0 \leq q < n} V_q^N(h_{q,n}^N) + \frac{1}{\sqrt{N}}\ R_n^N \tag{6.13}$$

with a second-order remainder term R_n^N such that

$$\mathbb{E}(|R_n^N|^m)^{1/m} \leq b(2m)^2 r(n)$$

for any $m \geq 1$, with some constant

$$r(n) \leq 8 \sum_{0 \leq p < n} g_p(2 g_p \tau_{1,1}(p)^2 + \tau_{3,1}(p)).$$

Proof:
Using the exponential formulae (6.12), we have

$$\begin{aligned}
\sqrt{N} \log \overline{\gamma}_n^N(\mathbb{1}) &= \sqrt{N} \log \left(1 + \frac{1}{\sqrt{N}} \overline{W}_n^{\gamma,N}(\mathbb{1})\right) \\
&= \sum_{0 \leq p < n} \int_0^1 \frac{W_p^{\eta,N}(\overline{G}_p)}{1 + \frac{t}{\sqrt{N}} W_p^{\eta,N}(\overline{G}_p)} dt.
\end{aligned}$$

This implies that

$$\sqrt{N} \log \overline{\gamma}_n^N(\mathbb{1}) = \sum_{0 \leq p < n} W_p^{\eta,N}(\overline{G}_p) + \frac{1}{\sqrt{N}} R_n^{N,1}$$

with the (negative) second order remainder term

$$R_n^{N,1} = -\sum_{0\le p<n} \int_0^1 t \frac{W_p^{\eta,N}(\overline{G}_p)^2}{1+\frac{t}{\sqrt{N}} W_p^{\eta,N}(\overline{G}_p)} dt.$$

On the other hand, using (6.11) we have

$$\sum_{0\le p<n} W_p^{\eta,N}(\overline{G}_p) = \sum_{0\le q<n} V_q^N(h_{q,n}^N) + \frac{1}{\sqrt{N}} R_n^{N,2}$$

with the second-order remainder term

$$R_n^{N,2} := -\sum_{0\le q\le p<n} \frac{1}{\eta_q^N(G_{q,p}^N)} V_q^N(G_{q,p}^N) V_q^N[d_{q,p}^N(\overline{G}_p)].$$

This gives the decomposition (6.13), with the second remainder order term

$$R_n^N := R_n^{N,1} + R_n^{N,2}.$$

Using the fact that

$$1 + \frac{t}{\sqrt{N}} W_p^{\eta,N}(\overline{G}_p) = t\eta_p^N(\overline{G}_p) + (1-t) \ge t g_p^-$$

for any $t \in]0,1]$, with $g_p^- := \inf_x \overline{G}_p(x)$, we find that

$$|R_n^{N,1}| \le \sum_{0\le p<n} \frac{1}{g_p^-} W_p^{\eta,N}(\overline{G}_p)^2.$$

Using (6.6), we prove that

$$\mathbb{E}(|R_n^{N,1}|^r)^{1/r} \le 4b(2r)^2 \sum_{0\le p<n} \frac{1}{g_p^-} \operatorname{osc}(\overline{G}_p)^2 \tau_{1,1}(p)^2$$

from which we conclude that

$$\mathbb{E}(|R_n^{N,1}|^r)^{1/r} \le (4b(2r))^2 \sum_{0\le p<n} g_p^2 \tau_{1,1}(p)^2.$$

In much the same way, we have

$$\mathbb{E}(|R_n^{N,2}|^r)^{1/r} \le \sum_{0\le q\le p<n} g_{q,p} \mathbb{E}(|V_q^N(G_{q,p}^N)|^{2r})^{1/2r} \mathbb{E}(|V_q^N[d_{q,p}^N(\overline{G}_p)]|^{2r})^{1/2r}$$

and using (6.4), we prove that

$$\mathbb{E}(|R_n^{N,2}|^r)^{1/r} \leq 8b(2r)^2 \sum_{0\leq q\leq p<n} g_p g_{q,p}^3 \beta(P_{q,p}).$$

This ends the proof of the lemma. ∎

We are now in position to state and to prove the following concentration theorem.

Theorem 6.13. For any $N \geq 1$, $\epsilon \in \{+1, -1\}$, $n \geq 0$, and for any

$$\varsigma_n^\star \geq \sup_{0\leq q\leq n} \varsigma_{q,n} \quad \text{with } \varsigma_{q,n} := \frac{4}{n} \sum_{q\leq p<n} g_{q,p} g_p \ \beta(P_{q,p})$$

the probability of the following events

$$\frac{\epsilon}{n} \log \overline{\gamma}_n^N(\mathbb{1}) \leq \frac{1}{N}\overline{r}(n)(1 + (L_0^\star)^{-1}(x)) + \varsigma_n^\star \ \overline{\sigma}_n^2 (L_1^\star)^{-1}\left(\frac{x}{N\overline{\sigma}_n^2}\right)$$

is greater than $1 - e^{-x}$, for any $x \geq 0$, with the parameters

$$\overline{\sigma}_n^2 := \sum_{0\leq q<n} \sigma_q^2 (\varsigma_{q,n}/\varsigma_n^\star)^2 \quad \text{and} \quad \overline{r}(n) = r(n)/n.$$

Before delving into the proof of the theorem, we present simple arguments to derive exponential concentration inequalities for the quantities $|\overline{\gamma}_n^N(\mathbb{1}) - 1|$. Suppose that for any $\epsilon \in \{+1, -1\}$, the probability of events

$$\frac{\epsilon}{n} \log \overline{\gamma}_n^N(\mathbb{1}) \leq \rho_n^N(x)$$

is greater than $1 - e^{-x}$, for any $x \geq 0$, for some function ρ_n^N such that

$$\rho_n^N(x) \to_{N\to\infty} 0.$$

In this case, the probability of event

$$-(1 - e^{-n\rho_n^N(x)}) \leq \overline{\gamma}_n^N(\mathbb{1}) - 1 \leq e^{n\rho_n^N(x)} - 1$$

is greater than $1 - 2e^{-x}$, for any $x \geq 0$. Choosing N large enough so that $\rho_n^N(x) \leq 1/n$ we have

$$-2n\rho_n^N(x) \leq -(1 - e^{-n\rho_n^N(x)}) \quad \text{and} \quad e^{n\rho_n^N(x)} - 1 \leq 2n\rho_n^N(x)$$

from which we conclude that the probability of event

$$\mathbb{P}(|\overline{\gamma}_n^N(\mathbb{1}) - 1| \leq 2n\ \rho_n^N(x)) \geq 1 - 2e^{-x}.$$

Now, we come to the proof of the theorem.

Proof of Theorem 6.13:
We use the same line of arguments as those used in subsection 6.4.1. First, we observe that

$$\begin{aligned}\|h_{q,n}^N\| &\leq \sum_{q \leq p < n} \|d_{q,p}^N(\overline{G}_p)\| \\ &\leq \sum_{q \leq p < n} g_{q,p}\mathrm{osc}(P_{q,p}(\overline{G}_p)) \leq 2 \sum_{q \leq p < n} g_{q,p}g_p\ \beta(P_{q,p}) = c_{q,n}/2\end{aligned}$$

and $\mathrm{osc}(h_{q,n}^N) \leq c_{q,n}$. Now, we use the following decompositions

$$\sum_{0 \leq q < n} V_q^N(h_{q,n}^N) = a_n^\star \sum_{0 \leq q < n} V_q^N(\delta_{q,n}^N)$$

with the $\mathcal{G}_{q-1}^N$-measurable functions

$$\delta_{q,n}^N = h_{q,n}^N / a_n^\star \in \mathrm{Osc}(E_q) \cap \mathcal{B}_1(E_q)$$

and for any constant $a_n^\star \geq \sup_{0 \leq q \leq n} c_{q,n}$.

On the other hand, we have the almost sure variance estimate

$$\mathbb{E}(V_q^N[\delta_{q,n}^N]^2 | \mathcal{G}_{q-1}^N) \leq \sigma_q^2 \mathrm{osc}(h_{q,n}^N)^2 / a_n^{\star 2} \leq \sigma_q^2 c_{q,n}^2 / a_n^{\star 2}$$

from which we conclude that

$$\mathbb{E}(V_q^N[\delta_{q,n}^N]^2) \leq \sigma_q^2 c_{q,n}^2 / a_n^{\star 2}.$$

This shows that the regularity condition stated in (5.2) is met by replacing the parameters σ_q in the variance formula (5.2) by the constants $\sigma_q c_{q,n}/a_n^\star$, with the uniform local variance parameters σ_p defined in (3.20).

The end of the proof is now a direct consequence of Theorem 5.2. This ends the proof of the theorem. ∎

Corollary 6.14. We assume that one of the regularity conditions $\mathbf{H_m(G,M)}$ stated in subsection 3.4.1 is met for some $m \geq 0$, and we set

$$p_m(x) := c_1(m)(1 + 2(x + \sqrt{x})) + c_2(m)x \quad \text{and} \quad q_m(x) = c_3(m)\sqrt{x}$$

with the parameters

$$\begin{aligned} c_1(m) &= (4g\overline{\tau}_{1,1}(m))^2 + 8g\overline{\tau}_{3,1}(m) \\ c_2(m) &= 4(\beta_m g^{m+1})/3 \quad \text{and} \quad c_3(m) = 4g\sqrt{2\overline{\tau}_{2,2}(m)\sigma^2}. \end{aligned}$$

In the above displayed formula, $\overline{\tau}_{2,2}(m)$ and $\overline{\kappa}(m)$ stands for the parameters defined in Corollary 3.5, and σ_n is the uniform local variance parameter defined in (3.20).

In this situation, for any $N \geq 1$, and any $\epsilon \in \{+1, -1\}$, the probability of each of the following events

$$\frac{\epsilon}{n} \log \overline{\gamma}_n^N(\mathbb{1}) \leq \frac{1}{N} p_m(x) + \frac{1}{\sqrt{N}} q_m(x)$$

is greater than $1 - e^{-x}$, for any $x \geq 0$.

Proof:

Under condition $\mathbf{H_m(G,M)}$, we have

$$r(n)/n \leq (4g\overline{\tau}_{1,1}(m))^2 + 8g\overline{\tau}_{3,1}(m)$$

and for any $p < n$

$$\begin{aligned} \varsigma_{p,n}^2 &= \left(\frac{4g}{n}\right)^2 (n-p)^2 \left(\frac{1}{n-p} \sum_{p \leq q < n} g_{p,q}\ \beta(P_{p,q})\right)^2 \\ &\leq \frac{(4g)^2}{n} \frac{(n-p)}{n} \sum_{p \leq q < n} g_{p,q}^2\ \beta(P_{p,q})^2. \end{aligned}$$

This implies that

$$\sum_{0 \leq p < n} \varsigma_{p,n}^2 \leq \frac{(4g)^2}{n} \sum_{0 \leq q < n} \tau_{2,2}(q) \leq (4g)^2 \overline{\tau}_{2,2}(m).$$

In much the same way, we prove that $\varsigma_n^\star \leq 4\beta_m g^{m+1}$. The end of the proof is now a consequence of the estimates (4.20) and (4.21). This ends the proof of the corollary. ■

6.6 Backward Particle Markov Models

This subsection is concerned with the concentration properties of the backward Markov particle measures defined in (1.18). Without further mention, we assume that the Markov transitions M_n satisfy the regularity condition (1.10), and we consider the random fields defined below.

Definition 6.4. We let $W_n^{\Gamma,N}$ and $W_n^{\mathbb{Q},N}$ be random field models defined by

$$W_n^{\Gamma,N} = \sqrt{N}(\Gamma_n^N - \Gamma_n) \quad \text{and} \quad W_n^{\mathbb{Q},N} = \sqrt{N}(\mathbb{Q}_n^N - \mathbb{Q}_n).$$

The analysis of the fluctuation of random fields of backward particle models is a little more involved than that of the genealogical tree particle models. The main difficulty is to deal with the nonlinear dependency of these backward particle Markov chain models with the flow of particle measures η_n^N.

In subsection 6.6.1, we provide some preliminary key backward conditioning principles. We also introduce some predictable integral operators involved in the first-order expansions of the fluctuation of random fields discussed in subsection 6.6.3. In subsection 6.6.2, we illustrate these models in the context of additive functional models. In subsection 6.6.4, we assemble the semigroup techniques developed earlier to derive a series of quantitative concentration inequalities.

6.6.1 Some Preliminary Conditioning Principles

By definition of the unnormalized Feynman-Kac measures Γ_n, we have

$$\Gamma_n(d(x_0,\ldots,x_n)) = \Gamma_p(d(x_0,\ldots,x_p))\Gamma_{n|p}(x_p, d(x_{p+1},\ldots,x_n))$$

with

$$\Gamma_{n|p}(x_p, d(x_{p+1},\ldots,x_n)) = \prod_{p<q\leq n} Q_q(x_{q-1}, dx_q).$$

This implies that

$$\mathbb{Q}_n(d(x_0,\ldots,x_n)) = \mathbb{Q}_{n,p}(d(x_0,\ldots,x_p)) \times \mathbb{Q}_{n|p}(x_p, d(x_{p+1},\ldots,x_n))$$

with the $\mathbb{Q}_n$-distribution of the random states $(X_0,\ldots,X_p)$

$$\mathbb{Q}_{n,p}(d(x_0,\ldots,x_p)) := \frac{1}{\eta_p(G_{p,n})}\mathbb{Q}_p(d(x_0,\ldots,x_p))G_{p,n}(x_p)$$

and the $\mathbb{Q}_n$-conditional distribution of $(X_{p+1},\ldots,X_n)$ given the random state $X_p = x_p$ defined by

$$\mathbb{Q}_{n|p}(x_p, d(x_{p+1},\ldots,x_n)) = \frac{1}{\Gamma_{n|p}(\mathbb{1})(x_p)}\Gamma_{n|p}(x_p, d(x_{p+1},\ldots,x_n)).$$

Now, we discuss some backward conditioning principles. Using the backward Markov chain formulation (1.11), we have

$$\mathbb{Q}_n(d(x_0,\ldots,x_n)) = \eta_n(dx_n)\mathbb{Q}_{n|n}(x_n, d(x_0,\ldots,x_{n-1}))$$

with the $\mathbb{Q}_n$-conditional distribution of $(X_0,\ldots,X_{n-1})$ given the terminal random state $X_n = x_n$ defined by the backward Markov transition

$$\mathbb{Q}_{n|n}(x_n, d(x_0,\ldots,x_{n-1})) := \prod_{q=1}^{n} \mathbb{M}_{q,\eta_{q-1}}(x_q, dx_{q-1}).$$

By construction, the $\mathbb{Q}_n^N$-conditional distribution of $(X_0,\ldots,X_{n-1})$ given the terminal random state $X_n = x_n$ is also defined by the particle backward Markov transition given by

$$\mathbb{Q}_{n|n}^N(x_n, d(x_0,\ldots,x_{n-1})) := \prod_{q=1}^{n} \mathbb{M}_{q,\eta_{q-1}^N}(x_q, dx_{q-1}).$$

We check this claim, using the fact that

$$\mathbb{Q}_n^N(d(x_0,\ldots,x_n)) = \eta_n^N(dx_n)\ \mathbb{Q}_{n|n}^N(x_n, d(x_0,\ldots,x_{n-1})).$$

Definition 6.5. For any $0 \leq p \leq n$ and $N \geq 1$, we denote by $D_{p,n}^N$ and $L_{p,n}^N$ the $\mathcal{G}_{p-1}^N$-measurable integral operators defined by

$$\begin{aligned} &D_{p,n}^N(x_p, d(y_0,\ldots,y_n)) \\ &\quad := \mathbb{Q}_{p|p}^N(x_p, d(y_0,\ldots,y_{p-1}))\delta_{x_p}(dy_p)\Gamma_{n|p}(x_p, d(y_{p+1},\ldots,y_n)) \end{aligned} \tag{6.14}$$

and

$$L^N_{p,n}(x_p, d(y_0, \ldots, y_n)) \\ := \mathbb{Q}^N_{p|p}(x_p, d(y_0, \ldots, y_{p-1}))\delta_{x_p}(dy_p)\mathbb{Q}_{n|p}(x_p, d(y_{p+1}, \ldots, y_n)). \quad (6.15)$$

For $p \in \{0, n\}$, we use the convention

$$D^N_{n,n}(x_n, d(y_0, \ldots, y_n)) = L^N_{n,n}(x_n, d(y_0, \ldots, y_n)) \\ = \mathbb{Q}^N_{n|n}(x_n, d(y_0, \ldots, y_{n-1}))\ \delta_{x_n}(dy_n)$$

and

$$D^N_{0,n}(x_0, d(y_0, \ldots, y_n)) = \delta_{x_0}(dy_0)\Gamma_{n|0}(x_0, d(y_1, \ldots, y_n)) \\ L^N_{0,n}(x_0, d(y_0, \ldots, y_n)) = \delta_{x_0}(dy_0)\mathbb{Q}_{n|0}(x_0, d(y_1, \ldots, y_n)).$$

The main reason for introducing these integral operators comes from the following integral transport properties.

Lemma 6.15. For any $0 \leq p \leq n$, and any $N \geq 1$, and any function $\mathbf{f}_n$ on the path space $\mathbf{E}_n$, we have the almost sure formulae

$$\eta^N_p D^N_{p,n} = \eta^N_p(G_p) \times \Phi_{p+1}(\eta^N_p) D^N_{p+1,n} \quad (6.16)$$

and

$$\frac{\eta^N_p D^N_{p,n}(\mathbf{f}_n)}{\eta^N_p D^N_{p,n}(\mathbb{1})} = \Psi_{G_{p,n}}(\eta^N_p) L^N_{p,n}(\mathbf{f}_n) \quad (6.17)$$

$$= \Psi_{G_{p+1,n}}(\Phi_{p+1}(\eta^N_p)) L^N_{p+1,n}(\mathbf{f}_n). \quad (6.18)$$

Proof. We check (6.16) using the fact that

$$\Gamma_{n|p}(x_p, d(y_{p+1}, \ldots, y_n)) \\ = Q_{p+1}(x_p, dy_{p+1})\Gamma_{n|p+1}(x_{p+1}, d(y_{p+2}, \ldots, y_n)) \quad (6.19)$$

and

$$\eta^N_p(dx_p)Q_{p+1}(x_p, dy_{p+1}) = \eta^N_p Q_{p+1}(dy_{p+1}) \times \mathbb{M}_{p+1,\eta^N_p}(y_{p+1}, dx_p). \quad (6.20)$$

More precisely, we have

$$\begin{aligned}\eta_p^N(dx_p)D_{p,n}^N(x_p,d(y_0,\ldots,y_n))&\\ :=\eta_p^N(dx_p)Q_{p+1}(x_p,dy_{p+1})\mathbb{Q}_{p|p}^N(x_p,d(y_0,\ldots,y_{p-1}))&\\ \times\delta_{x_p}(dy_p)\Gamma_{n|p+1}(y_{p+1},d(y_{p+2},\ldots,y_n)).&\end{aligned}$$

Using (6.20), this implies that

$$\begin{aligned}\eta_p^N(dx_p)D_{p,n}^N(x_p,d(y_0,\ldots,y_n))&\\ :=\eta_p^NQ_{p+1}(dy_{p+1})\mathbb{M}_{p+1,\eta_p^N}(y_{p+1},dx_p)\mathbb{Q}_{p|p}^N(x_p,d(y_0,\ldots,y_{p-1}))&\\ \times\delta_{x_p}(dy_p)\Gamma_{n|p+1}(y_{p+1},d(y_{p+2},\ldots,y_n))&\end{aligned}$$

from which we conclude that

$$\begin{aligned}\eta_p^ND_{p,n}^N(\mathbf{f_n})&\\ :=\int\eta_p^NQ_{p+1}(dy_{p+1})\mathbb{Q}_{p+1|p+1}^N(y_{p+1},d(y_0,\ldots,y_p))&\\ \times\Gamma_{n|p+1}(y_{p+1},d(y_{p+2},\ldots,y_n))\mathbf{f_n}(y_0,\ldots,y_n)&\\ =(\eta_p^NQ_{p+1})D_{p+1,n}^N(\mathbf{f_n}).&\end{aligned}$$

This ends the proof of the first assertion. Now, using (6.16) we have

$$\frac{\eta_p^ND_{p,n}^N(\mathbf{f_n})}{\eta_p^ND_{p,n}^N(\mathbb{1})}=\frac{\Phi_{p+1}(\eta_p^N)D_{p+1,n}^N(\mathbf{f_n})}{\Phi_{p+1}(\eta_p^N)D_{p+1,n}^N(\mathbb{1})}.$$

Recalling that

$$D_{p,n}^N(\mathbb{1})=Q_{p,n}(\mathbb{1})=G_{p,n}$$

we readily prove (6.17) and (6.18). This ends the proof of the lemma. ∎

6.6.2 Additive Functional Models

Here, we provide a brief discussion on the action of the operators $D_{p,n}^N$ and $L_{p,n}^N$ on additive linear functionals

$$\mathbf{f_n}(x_0,\ldots,x_n)=\sum_{p=0}^n f_p(x_p) \tag{6.21}$$

associated with some collection of functions $f_n \in \mathrm{Osc}(E_n)$.

$$D_{p,n}^N(\mathbf{f_n}) = Q_{p,n}(1)\left[\sum_{0\leq q<p}[\mathbb{M}_{p,\eta_{p-1}^N}\ldots\mathbb{M}_{q+1,\eta_q^N}](f_q) + \sum_{p\leq q\leq n} R_{p,q}^{(n)}(f_q)\right]$$

with a triangular array of Markov transitions $R_{p,q}^{(n)}$ introduced in Definition 3.2. By definition of $L_{p,n}^N$, we also have that

$$L_{p,n}^N(\mathbf{f_n}) = \sum_{0\leq q<p}[\mathbb{M}_{p,\eta_{p-1}^N}\ldots\mathbb{M}_{q+1,\eta_q^N}](f_q) + \sum_{p\leq q\leq n} R_{p,q}^{(n)}(f_q).$$

Using the estimates (3.9), we prove the following upper bounds

$$\begin{aligned}\mathrm{osc}(L_{p,n}^N(\mathbf{f_n})) \leq & \sum_{0\leq q<p}\beta(\mathbb{M}_{p,\eta_{p-1}^N}\ldots\mathbb{M}_{q+1,\eta_q^N}) \\ & + \sum_{p\leq q\leq n} g_{q,n}\times\beta(P_{p,q}).\end{aligned}$$

There are many ways to control the Dobrushin operator norm of the product of the random matrices defined in (1.19). For instance, we can use the multiplicative formulae

$$\beta(\mathbb{M}_{p,\eta_{p-1}^N}\ldots\mathbb{M}_{q+1,\eta_q^N}) \leq \prod_{p<k\leq q}\beta(\mathbb{M}_{k+1,\eta_k^N}).$$

One of the simplest ways to proceed, is to assume that

$$H_n(x,y) \leq \tau\ H_n(x,y') \tag{6.22}$$

for any x, y, y', and for some finite constant $\tau < \infty$. In this situation, we find that

$$\mathbb{M}_{k+1,\eta_k^N}(y,dx) \leq \tau^2\ \mathbb{M}_{k+1,\eta_k^N}(y',dx)$$

from which we conclude that

$$\beta(\mathbb{M}_{k+1,\eta_k^N}) \leq 1-\tau^{-2}.$$

We further assume that the condition $\mathbf{H_m}(\mathbf{G},\mathbf{M})$ stated in subsection 3.4.1 is met for some $m \geq 1$. In this situation, we have

$$\begin{aligned}&\mathrm{osc}(L_{p,n}^N(\mathbf{f_n})) \\ &\quad\leq \sum_{0\leq q<p}(1-\tau^{-2})^{(p-q)} + \chi_m g^m \sum_{p\leq q\leq n}(1-g^{-(m-1)}\chi_m^{-2})^{\lfloor (q-p)/m\rfloor}\end{aligned}$$

from which we prove the following uniform estimates

$$\sup_{0\leq p\leq n} \mathrm{osc}(L_{p,n}^N(\mathbf{f_n})) \leq \tau^2 + m\ g^{2m-1}\chi_m^3. \tag{6.23}$$

6.6.3 A Stochastic Perturbation Analysis

As in subsection 6.3, we develop a stochastic perturbation analysis that allows us to express $W_n^{\Gamma,N}$ and $W_n^{\mathbb{Q},N}$ in terms of the local sampling random fields $(V_p^N)_{0\leq p\leq n}$.

These first-order expansions presented will be expressed in terms of the first-order functions $d_\nu\Psi_G(f)$ introduced in (6.2), and the random $\mathcal{G}_{p-1}^N$-measurable functions $G_{p,n}^N$ introduced in Definition 6.1.

Definition 6.6. For any $N \geq 1$, any $0 \leq p \leq n$, and any function $\mathbf{f_n}$ on the path space $\mathbf{E_n}$, we let $\mathbf{d_{p,n}^N}(\mathbf{f_n})$ be the $\mathcal{G}_{p-1}^N$-measurable functions

$$\mathbf{d_{p,n}^N}(\mathbf{f_n}) = d_{\Phi_p(\eta_{p-1}^N)}\Psi_{G_{p,n}}(L_{p,n}^N(\mathbf{f_n})).$$

We are now in a position to state and to prove the following decomposition theorem.

Theorem 6.16. For any $0 \leq p \leq n$, and any function $\mathbf{f_n}$ on the path space E^{n+1}, we have

$$\mathbb{E}(\Gamma_n^N(\mathbf{f_n})|\mathcal{G}_p^N) = \gamma_p^N(D_{p,n}^N(\mathbf{f_n})). \tag{6.24}$$

In addition, we have

$$W_n^{\Gamma,N}(\mathbf{f_n}) = \sum_{p=0}^n \gamma_p^N(1)V_p^N(D_{p,n}^N(\mathbf{f_n})) \tag{6.25}$$

$$W_n^{\mathbb{Q},N}(\mathbf{f_n}) = \sum_{p=0}^n \frac{1}{\eta_p^N(G_{p,n}^N)}V_p^N(\mathbf{d_{p,n}^N}(\mathbf{f_n})) \tag{6.26}$$

$$= \sum_{p=0}^n V_p^N(\mathbf{d_{p,n}^N}(\mathbf{f_n})) - \sum_{p=0}^n \frac{1}{\eta_p^N(G_{p,n}^N)}\frac{1}{\sqrt{N}}V_p^N(G_{p,n}^N) \times V_p^N(\mathbf{d_{p,n}^N}(\mathbf{f_n})). \tag{6.27}$$

Proof:
To prove the first assertion, we use a backward induction on the parameter p. For $p = n$, the result is immediate since we have

$$\Gamma_n^N(\mathbf{f_n}) = \gamma_n^N(\mathbb{1})\ \eta_n^N(D_{n,n}^N(\mathbf{f_n})).$$

We suppose that the formula is valid at a given rank $p \leq n$. In this situation, using the fact that $D_{p,n}^N(\mathbf{f_n})$ is a $\mathcal{G}_{p-1}^N$-measurable function, we prove that

$$\mathbb{E}(\Gamma_n^N(\mathbf{f_n})|\mathcal{G}_{p-1}^N) = \mathbb{E}(\gamma_p^N(D_{p,n}^N(\mathbf{f_n}))|\mathcal{G}_{p-1}^N) = (\gamma_{p-1}^N Q_p)D_{p,n}^N(\mathbf{f_n}) \qquad (6.28)$$

Applying (6.16), we also have that

$$\gamma_{p-1}^N Q_p D_{p,n}^N = \gamma_{p-1}^N D_{p-1,n}^N$$

from which we conclude that the desired formula is satisfied at rank $(p-1)$. This ends the proof of the first assertion.

Now, combining Lemmas 6.15 and (6.24), the proof of the second assertion is simply based on the following decomposition

$$\begin{aligned}
&(\Gamma_n^N - \Gamma_n)(\mathbf{f_n}) \\
&\quad = \sum_{p=0}^{n}[\mathbb{E}(\Gamma_n^N(\mathbf{f_n})|\mathcal{G}_p^N) - \mathbb{E}(\Gamma_n^N(\mathbf{f_n})|\mathcal{G}_{p-1}^N)] \\
&\quad = \sum_{p=0}^{n}\gamma_p^N(1)\left(\eta_p^N(D_{p,n}^N(\mathbf{f_n})) - \frac{1}{\eta_{p-1}^N(G_{p-1})}\eta_{p-1}^N(D_{p-1,n}^N(\mathbf{f_n}))\right).
\end{aligned}$$

To prove the final decomposition, we use the fact that

$$[\mathbb{Q}_n^N - \mathbb{Q}_n](\mathbf{f_n}) = \sum_{0\leq p\leq n}\left(\frac{\eta_p^N D_{p,n}^N(\mathbf{f_n})}{\eta_p^N D_{p,n}^N(1)} - \frac{\eta_{p-1}^N D_{p-1,n}^N(\mathbf{f_n})}{\eta_{p-1}^N D_{p-1,n}^N(1)}\right)$$

with the conventions $\eta_{-1}^N D_{-1,n}^N = \eta_0\Gamma_{n|0}$, for $p = 0$.

Finally, we use (6.17) and (6.18) to check that

$$[\mathbb{Q}_n^N - \mathbb{Q}_n] = \sum_{0\leq p\leq n}(\Psi_{G_{p,n}}(\eta_p^N) - \Psi_{G_{p,n}}(\Phi_p(\eta_{p-1}^N)))L_{p,n}^N.$$

We end the proof using the first-order expansions of the Boltzmann-Gibbs transformation developed in subsection 6.1:

$$\begin{aligned}\sqrt{N}(\Psi_{G_{p,n}}(\eta_p^N) - \Psi_{G_{p,n}}(\Phi_p(\eta_{p-1}^N)))L_{p,n}^N(\mathbf{f_n}) \\ = \frac{1}{\eta_p^N(G_{p,n}^N)} V_p^N(\mathbf{d}_{\mathbf{p,n}}^{\mathbf{N}}(\mathbf{f_n})) \\ = V_p^N(\mathbf{d}_{\mathbf{p,n}}^{\mathbf{N}}(\mathbf{f_n})) - \frac{1}{\eta_p^N(G_{p,n}^N)} \frac{1}{\sqrt{N}} V_p^N(G_{p,n}^N) \times V_p^N(\mathbf{d}_{\mathbf{p,n}}^{\mathbf{N}}(\mathbf{f_n})).\end{aligned}$$

This ends the proof of the theorem. ∎

6.6.4 Concentration Inequalities

6.6.4.1 Finite Marginal Models

Given a bounded function $\mathbf{f_n}$ on the path space $\mathbf{E_n}$, we further assume that we have some almost sure estimate

$$\sup_{N\geq 1} \operatorname{osc}(L_{p,n}^N(\mathbf{f_n})) \leq l_{p,n}(\mathbf{f_n}) \tag{6.29}$$

for some finite constant $l_{p,n}(\mathbf{f_n})$ $(\leq \|\mathbf{f_n}\|)$. For instance, for additive functionals of the form (6.21), we have proved in subsection 6.6.2 the following uniform estimates

$$\operatorname{osc}(L_{p,n}^N(\mathbf{f_n})) \leq \tau^2 + m\ g^{2m-1}\chi_m^3$$

which are valid for any $N \geq 1$ and any $0 \leq p \leq n$, as soon as the mixing condition $\mathbf{H_m}(\mathbf{G}, \mathbf{M})$ stated in subsection 3.4.1 is met for some $m \geq 1$, and the regularity (6.22) is satisfied for some finite τ.

For any additive functional $\mathbf{f_n}$ of the form (6.21), we denote by $\overline{\mathbf{f}}_n = \mathbf{f_n}/(n+1)$ the normalized additive functional.

Lemma 6.17. For any $N \geq 1$, $n \geq 0$, and any bounded function $\mathbf{f_n}$ on the path space $\mathbf{E_n}$, we have the first-order decomposition

$$W_n^{\mathbb{Q},N}(\mathbf{f_n}) = \sum_{p=0}^{n} V_p^N(\mathbf{d}_{\mathbf{p,n}}^{\mathbf{N}}(\mathbf{f_n})) + \frac{1}{\sqrt{N}} R_n^N(\mathbf{f_n}) \tag{6.30}$$

with a second-order remainder term $R_n^N(\mathbf{f_n})$ such that

$$\mathbb{E}(|R_n^N(\mathbf{f_n})|^m)^{1/m} \leq b(2m)^2\ r_n(\mathbf{f_n})$$

for any $m \geq 1$, with some finite constant

$$r_n(\mathbf{f_n}) \leq 4 \sum_{0 \leq p \leq n} g_{p,n}^2 l_{p,n}(\mathbf{f_n}). \tag{6.31}$$

Proof. First, we notice that

$$\|\mathbf{d}^{\mathbf{N}}_{\mathbf{p},\mathbf{n}}(\mathbf{f_n})\| \leq \frac{\|G_{p,n}\|}{\Phi_p(\eta^N_{p-1})(G_{p,n})} \mathrm{osc}(L^N_{p,n}(\mathbf{f_n})) \leq g_{p,n} \mathrm{osc}(L^N_{p,n}(\mathbf{f_n})).$$

Using (6.27), we find the decomposition (6.30) with the second-order remainder term

$$R^N_n(\mathbf{f_n}) = -\sum_{p=0}^{n} R^N_{p,n}(\mathbf{f_n})$$

with

$$R^N_{p,n}(\mathbf{f_n}) := \frac{1}{\eta^N_p(G^N_{p,n})} \; V^N_p(G^N_{p,n}) \; V^N_p(\mathbf{d}^{\mathbf{N}}_{\mathbf{p},\mathbf{n}}(\mathbf{f_n})).$$

On the other hand, we have

$$\begin{aligned}
&\mathbb{E}(|R^N_{p,n}(\mathbf{f_n})|^m | \mathcal{G}^N_{p-1})^{1/m} \\
&\quad \leq g_{p,n} \mathbb{E}(|V^N_p(G_{p,n}/\|G_{p,n}\|)|^{2m} | \mathcal{G}^N_{p-1})^{1/(2m)} \\
&\qquad \times \mathbb{E}(|V^N_p(d^N_{p,n}(\mathbf{f_n}))|^{2m} | \mathcal{G}^N_{p-1})^{1/(2m)}.
\end{aligned}$$

Using (6.4), we prove that

$$\mathbb{E}(|R^N_{p,n}(\mathbf{f_n})|^m | \mathcal{G}^N_{p-1})^{1/m} \leq 4b(2m)^2 \; g^2_{p,n} \; l_{p,n}(\mathbf{f_n}).$$

The end of the proof is now clear. This ends the proof of the lemma. ∎

Theorem 6.18. For any $N \geq 1$, $n \geq 0$, and any bounded function $\mathbf{f_n}$ on the path space $\mathbf{E_n}$, the probability of the events

$$[\mathbb{Q}^N_n - \mathbb{Q}_n](\mathbf{f_n}) \leq \frac{r_n(\mathbf{f_n})}{N}(1 + (L^\star_0)^{-1}(x)) + 2b_n \; \overline{\sigma}^2_n (L^\star_1)^{-1}\left(\frac{x}{N\overline{\sigma}^2_n}\right)$$

is greater than $1 - e^{-x}$, for any $x \geq 0$, with

$$\overline{\sigma}_n^2 := \frac{1}{b_n^2} \sum_{0 \leq p \leq n} g_{p,n}^2 l_{p,n}(\mathbf{f_n})^2 \sigma_p^2$$

and for any choice of $b_n \geq \sup_{0 \leq p \leq n} g_{p,n} l_{p,n}(\mathbf{f_n})$. In the above displayed formulae, σ_n are the uniform local variance parameters defined in (3.20), $l_{p,n}(\mathbf{f_n})$ and $r_n(\mathbf{f_n})$ are the parameters defined in (6.29) and (6.31), respectively.

Before proceeding into the proof of the theorem, we present some direct consequences of these concentration inequalities for normalized additive functionals (we use the estimates (4.20) and (4.21)).

Corollary 6.19. We assume that the mixing condition $\mathbf{H_m}(\mathbf{G}, \mathbf{M})$ stated in subsection 3.4.1 is met for some $m \geq 1$, and the regularity (6.22) is satisfied for some finite τ. We also suppose that the parameters σ_n defined in (3.20) are uniformly bounded $\sigma = \sup_{n \geq 0} \sigma_n < \infty$ and we set

$$c_1(m) := 2 g^m \chi_m \ (\tau^2 + m \ g^{2m-1} \chi_m^3) \quad \text{and} \quad c_2(m) := 2 (g^m \chi_m) c_1(m).$$

In this notation, for any $N \geq 1$, $n \geq 0$, and any normalized additive functional $\overline{\mathbf{f}}_n$ on the path space $\mathbf{E_n}$, the probability of the events

$$[\mathbb{Q}_n^N - \mathbb{Q}_n](\overline{\mathbf{f}}_n)$$
$$\leq \frac{c_2(m)}{N} \ (1 + (L_0^\star)^{-1}(x)) + c_1(m) \sigma^2 \ (L_1^\star)^{-1} \left(\frac{x}{N(n+1)\sigma^2} \right)$$

is greater than $1 - e^{-x}$, for any $x \geq 0$.

Corollary 6.20. We assume that the assumptions of corollary 6.19 are satisfied, and we set

$$p_{m,n}(x) = c_2(m)(1 + 2(x + \sqrt{x})) + \frac{c_1(m)}{3(n+1)} x$$

and

$$q_{m,n}(x) = c_1(m)\sqrt{\frac{2x\sigma^2}{(n+1)}}$$

with the constants $c_1(m)$ and $c_2(m)$ defined in Corollary 6.19.

In this situation, the probability of the events

$$[\mathbb{Q}_n^N - \mathbb{Q}_n](\bar{\mathbf{f}}_n) \leq \frac{1}{N}\ p_{m,n}(x) + \frac{1}{\sqrt{N}} q_{m,n}(x)$$

is greater than $1 - e^{-x}$, for any $x \geq 0$.

Proof of Theorem 6.18:
We use the same line of arguments as those used in subsection 6.4.1. Firstly, we notice that

$$\|\mathbf{d}^{\mathbf{N}}_{\mathbf{p},\mathbf{n}}(\mathbf{f_n})\| \leq g_{p,n}\ l_{p,n}(\mathbf{f_n}).$$

This yields the decompositions

$$\sum_{p=0}^{n} V_p^N(\mathbf{d}^{\mathbf{N}}_{\mathbf{p},\mathbf{n}}(\mathbf{f_n})) = \sum_{p=0}^{n} a_p V_p^N(\delta^{\mathbf{N}}_{\mathbf{p},\mathbf{n}}(\mathbf{f_n}))$$

with the functions

$$\delta^{\mathbf{N}}_{\mathbf{p},\mathbf{n}}(\mathbf{f_n}) = \mathbf{d}^{\mathbf{N}}_{\mathbf{p},\mathbf{n}}(\mathbf{f_n})/a_p \in \mathrm{Osc}(E_p) \cap \mathcal{B}_1(E_p)$$

and for any finite constants

$$a_p \geq 2 \sup_{0\leq p\leq n} g_{p,n}\ l_{p,n}(\mathbf{f_n}).$$

On the other hand, we also have that

$$\mathbb{E}(V_p^N(\delta^{\mathbf{N}}_{\mathbf{p},\mathbf{n}}(\mathbf{f_n}))^2 | \mathcal{G}_{p-1}^N) \leq \frac{4}{(a_n^\star)^2}\ \sigma_p^2\ g_{p,n}^2\ l_{p,n}(\mathbf{f_n})^2$$

with $a^\star := \sup_{0\leq p\leq n} a_p$, and the uniform local variance parameters σ_p^2 defined in (3.20).

This shows that the regularity condition stated in (5.2) is met by replacing the parameters σ_p in the variance formula (5.2) by the constants $2\sigma_p g_{p,n} l_{p,n}(\mathbf{f_n})/a_n^\star$, with the uniform local variance parameters σ_p defined in (3.20).

Using theorem 5.2, we easily prove the desired concentration property. This ends the proof of the theorem. ■

6.6.4.2 Empirical Processes

Using the same line of arguments as those used in subsection 6.4.2, we prove the following concentration inequality.

Theorem 6.21. We let $\mathcal{F}_n$ be a separable collection of measurable functions $\mathbf{f_n}$ on $\mathbf{E_n}$, such that $\|\mathbf{f_n}\| \leq 1$, $\mathrm{osc}(\mathbf{f_n}) \leq 1$, with finite entropy $I(\mathcal{F}_n) < \infty$.

$$\pi_\psi(\|W_n^{\mathbb{Q},N}\|_{\mathcal{F}_n}) \leq c_{\mathcal{F}_n} \sum_{p=0}^{n} g_{p,n} \|l_{p,n}\|_{\mathcal{F}_n}$$

with the functional $\mathbf{f_n} \in \mathcal{F}_n \mapsto l_{p,n}(\mathbf{f_n})$ defined in (6.29) and

$$c_{\mathcal{F}_n} \leq 24^2 \int_0^1 \sqrt{\log\left(8 + \mathcal{N}(\mathcal{F}_n, \epsilon)^2\right)} d\epsilon.$$

In particular, for any $n \geq 0$, and any $N \geq 1$, the probability of the following event

$$\sup_{\mathbf{f_n} \in \mathcal{F}_n} |\mathbb{Q}_n^N(\mathbf{f_n}) - \mathbb{Q}_n(\mathbf{f_n})| \leq \frac{c_{\mathcal{F}_n}}{\sqrt{N}} \sum_{p=0}^{n} g_{p,n} \ \|l_{p,n}\|_{\mathcal{F}_n} \ \sqrt{x + \log 2}$$

is greater than $1 - e^{-x}$, for any $x \geq 0$.

Proof:
Using (6.26), for any function $\mathbf{f_n} \in \mathrm{Osc}(\mathbf{E_n})$ we have the estimate

$$|W_n^{\mathbb{Q},N}(\mathbf{f_n})| \leq 2 \sum_{p=0}^{n} g_{p,n} \ l_{p,n}(\mathbf{f_n}) |V_p^N(\delta_{p,n}^N(f_n))|$$

with the $\mathcal{G}_{p-1}^N$-measurable random functions $\delta_{p,n}^N(\mathbf{f_n})$ on E_p defined by

$$\delta_{p,n}^N(\mathbf{f_n}) = \frac{1}{2 l_{p,n}(\mathbf{f_n})} \frac{G_{p,n}}{\|G_{p,n}\|} [L_{p,n}^N(\mathbf{f_n}) - \Psi_{G_{p,n}}(\Phi_p(\eta_{p-1}^N)) L_{p,n}^N(\mathbf{f_n})].$$

By construction, we have

$$\|\delta_{p,n}^N(\mathbf{f_n})\| \leq 1/2 \quad \text{and} \quad \delta_{p,n}^N(\mathbf{f_n}) \in \mathrm{Osc}(E_p).$$

Using the uniform estimate (6.3), if we set

$$\mathcal{G}_{p,n}^N := \delta_{p,n}^N(\mathcal{F}_n) = \{\delta_{p,n}^N(\mathbf{f_n}) : \mathbf{f_n} \in \mathcal{F}_n\}$$

then we also prove the almost sure upper bound

$$\sup_{N\geq 1}\mathcal{N}[\mathcal{G}_{p,n}^N,\epsilon]\leq\mathcal{N}(\mathcal{F}_n,\epsilon/2).$$

The end of the proof is now a direct consequence of Theorem 5.3, and Theorem 4.6. This ends the proof of the theorem. ∎

Definition 6.7. We let $\mathcal{F}=(\mathcal{F}_n)_{n\geq 0}$, be a sequence of separable collections $\mathcal{F}_n$ of measurable functions f_n on E_n, such that $\|f_n\|\leq 1$, $\mathrm{osc}(f_n)\leq 1$, and finite entropy $I(\mathcal{F}_n)<\infty$.

For any $n\geq 0$, we set

$$J_n(\mathcal{F}):=24^2\sup_{0\leq q\leq n}\int_0^1\sqrt{\log\left(8+\mathcal{N}(\mathcal{F}_q,\epsilon)^2\right)}\,d\epsilon.$$

We also denote by $\Sigma_n(\mathcal{F})$ the collection of additive functionnals defined by

$$\Sigma_n(\mathcal{F})=\left\{\begin{array}{l}\mathbf{f}_n\in\mathcal{B}(\mathbf{E}_n)\text{ such that }\forall\mathbf{x_n}=(x_0,\ldots,x_n)\in\mathbf{E_n}\\ \\ \qquad\mathbf{f_n}(\mathbf{x_n})=\displaystyle\sum_{p=0}^n f_p(x_p)\text{ with }f_p\in\mathcal{F}_p,\text{ for }0\leq p\leq n\end{array}\right\}$$

and

$$\overline{\Sigma}_n(\mathcal{F})=\{\mathbf{f}_n/(n+1)\ :\ \mathbf{f}_n\in\Sigma_n(\mathcal{F})\}.$$

Theorem 6.22. For any $N\geq 1$, and any $n\geq 0$, we have

$$\pi_\psi(\|W_n^{\mathbb{Q},N}\|_{\Sigma_n(\mathcal{F})})\leq a_n J_n(\mathcal{F})\sum_{p=0}^n g_{p,n}$$

with some constant

$$a_n\leq\sum_{0\leq q<p}\beta_{p,q}+\sum_{p\leq q\leq n}\beta(R_{p,q}^{(n)})$$

and for any a collection of $[0,1]$-valued parameters $\beta_{p,q}$ such that

$$\forall 0\leq q\leq p\quad\sup_{N\geq 1}\beta(\mathbb{M}_{p,\eta_{p-1}^N}\ldots\mathbb{M}_{q+1,\eta_q^N})\leq\beta_{p,q}.$$

Proof. By definition of the operator $\mathbf{d}^{\mathbf{N}}_{\mathbf{p},\mathbf{n}}$ given in definition 6.6, for additive functionals $\mathbf{f_n}$ of the form (6.21), we find that

$$\begin{aligned}\mathbf{d}'^{\mathbf{N}}_{\mathbf{p},\mathbf{n}}(\mathbf{f_n}) &:= \Phi_p(\eta^N_{p-1})(G_{p,n}) \times \mathbf{d}^{\mathbf{N}}_{\mathbf{p},\mathbf{n}}(\mathbf{f_n}) \\ &= G_{p,n}[Id - \Psi_{G_{p,n}}(\Phi_p(\eta^N_{p-1}))]L^N_{p,n}(\mathbf{f_n}) \\ &= \sum_{0\leq q<p} \mathbf{d}^{(\mathbf{N},\mathbf{1})}_{\mathbf{p},\mathbf{q},\mathbf{n}}(f_q) + \sum_{p\leq q\leq n} \mathbf{d}^{(\mathbf{N},\mathbf{2})}_{\mathbf{p},\mathbf{q},\mathbf{n}}(f_q)\end{aligned}$$

with

$$\begin{aligned}\mathbf{d}^{(\mathbf{N},\mathbf{1})}_{\mathbf{p},\mathbf{q},\mathbf{n}}(f_q) &= G_{p,n}[Id - \Psi_{G_{p,n}}(\Phi_p(\eta^N_{p-1}))](\mathbb{M}^{(N)}_{p,q}(f_q)) \\ \mathbf{d}^{(\mathbf{N},\mathbf{2})}_{\mathbf{p},\mathbf{q},\mathbf{n}}(f_q) &= G_{p,n}[Id - \Psi_{G_{p,n}}(\Phi_p(\eta^N_{p-1}))](R^{(n)}_{p,q}(f_q))\end{aligned}$$

and

$$\mathbb{M}^{(N)}_{p,q} := \mathbb{M}_{p,\eta^N_{p-1}} \ldots \mathbb{M}_{q+1,\eta^N_q}.$$

This implies that

$$\begin{aligned}|V^N_p(\mathbf{d}'^{\mathbf{N}}_{\mathbf{p},\mathbf{n}}(\mathbf{f_n}))| \leq 2\|G_{p,n}\| \Bigg(&\sum_{0\leq q<p} \beta_{p,q} V^N_p(\delta^{(\mathbf{N},\mathbf{1})}_{\mathbf{p},\mathbf{q},\mathbf{n}}(f_q)) \\ &+ \sum_{p\leq q\leq n} \beta(R^{(n)}_{p,q}) V^N_p(\delta^{(\mathbf{N},\mathbf{2})}_{\mathbf{p},\mathbf{q},\mathbf{n}}(f_q)) \Bigg) \qquad (6.32)\end{aligned}$$

with

$$\delta^{(\mathbf{N},\mathbf{1})}_{\mathbf{p},\mathbf{q},\mathbf{n}}(f_q) = \frac{1}{2\beta(\mathbb{M}^{(N)}_{p,q})} \frac{G_{p,n}}{\|G_{p,n}\|} G_{p,n}[Id - \Psi_{G_{p,n}}(\Phi_p(\eta^N_{p-1}))]\mathbb{M}^{(N)}_{p,q}(f_q)$$

and

$$\delta^{(\mathbf{N},\mathbf{2})}_{\mathbf{p},\mathbf{q},\mathbf{n}}(f_q) = \frac{1}{2\beta(R^{(n)}_{p,q})} \frac{G_{p,n}}{\|G_{p,n}\|} G_{p,n}[Id - \Psi_{G_{p,n}}(\Phi_p(\eta^N_{p-1}))]R^{(n)}_{p,q}(f_q).$$

By construction, we have that

$$\|\delta^{(\mathbf{N},\mathbf{i})}_{\mathbf{p},\mathbf{q},\mathbf{n}}(f_q)\| \leq 1/2 \quad \text{and} \quad \operatorname{osc}(\delta^{(\mathbf{N},\mathbf{i})}_{\mathbf{p},\mathbf{q},\mathbf{n}}(f_q)) \leq 1$$

for any $i \in \{1,2\}$. We set

$$\mathcal{G}^{(N,i)}_{p,q,n} := \delta^{(\mathbf{N},\mathbf{i})}_{\mathbf{p},\mathbf{q},\mathbf{n}}(\mathcal{F}_q) = \{\delta^{(\mathbf{N},\mathbf{i})}_{\mathbf{p},\mathbf{q},\mathbf{n}}(f_q) : f_q \in \mathcal{F}_q\}.$$

Using the uniform estimate (6.3), we also prove the almost sure upper bound

$$\sup_{N\geq 1} \mathcal{N}[\mathcal{G}_{p,q,n}^{(N,i)}, \epsilon] \leq \mathcal{N}(\mathcal{F}_q, \epsilon/2).$$

Using Theorem 5.3, we prove that

$$\pi_\psi(\sup_{\mathbf{f}_n \in \Sigma_n(\mathcal{F})} |V_p^N(\mathbf{d}'^{\mathbf{N}}_{\mathbf{p},\mathbf{n}}(\mathbf{f_n}))|) \leq a_n \|G_{p,n}\| J_n(\mathcal{F}).$$

The end of the proof of the theorem is now a direct consequence of the decomposition (6.26). This ends the proof of the theorem. ∎

Corollary 6.23. We further assume that the condition $\mathbf{H_m}(\mathbf{G},\mathbf{M})$ stated in subsection 3.4.1 is met for some $m \geq 1$, the condition (6.22) is satisfied for some τ, and we have $J(\mathcal{F}) := \sup_{n\geq 0} J_n(\mathcal{F}) < \infty$. In this situation, we have the uniform estimates

$$\sup_{n\geq 0} \pi_\psi(\|W_n^{\mathbb{Q},N}\|_{\overline{\Sigma}_n(\mathcal{F})}) \leq c_{\mathcal{F}}(m)$$

with some constant

$$c_{\mathcal{F}}(m) \leq \chi_m g^m (\tau^2 + m\ g^{2m-1}\chi_m^3) J(\mathcal{F}).$$

In particular, for any time horizon $n \geq 0$, and any $N \geq 1$, the probability of the following event

$$\|\mathbb{Q}_n^N - \mathbb{Q}_n\|_{\overline{\Sigma}_n(\mathcal{F})} \leq \frac{1}{\sqrt{N}} c_{\mathcal{F}}(m) \sqrt{x + \log 2}$$

is greater than $1 - e^{-x}$, for any $x \geq 0$.

Proof. When the conditions $\mathbf{H_m}(\mathbf{G},\mathbf{M})$ and (6.22) are satisfied, we proved in subsection 6.6.2 that

$$\sum_{0\leq q<p} \beta_{p,q} + \sum_{p\leq q\leq n} \beta(R_{p,q}^{(n)}) \leq \tau^2 + m g^{2m-1}\chi_m^3.$$

We end the proof of the theorem, recalling that $g_{p,n} \leq \chi_m g^m$. This ends the proof of the theorem. ∎

We end this section with a direct consequence of (4.10).

Corollary 6.24. We further assume that the condition $\mathbf{H_m}(\mathbf{G},\mathbf{M})$ stated in subsection 3.4.1 is met for some $m \geq 1$, the condition (6.22) is satisfied for some τ. We let $\mathcal{F}_n$ be the set of product functions of cell indicators in the path space $E_n = \mathbb{R}^d$, for some $d \geq 1$, $p \geq 0$.

In this situation, for any time horizon $n \geq 0$, and any $N \geq 1$, the probability of the following event

$$\|\mathbb{Q}_n^N - \mathbb{Q}_n\|_{\overline{\Sigma}_n(\mathcal{F})} \leq c(m)\sqrt{\frac{d}{N}(x+1)}$$

is greater than $1 - e^{-x}$, for any $x \geq 0$, with some constant

$$c(m) \leq c\ \chi_m g^m\ (\tau^2 + m\ g^{2m-1}\chi_m^3).$$

In the above display, c stands for some finite universal constant.

References

[1] R. Adamczak, "A tail inequality for suprema of unbounded empirical processes with applications to Markov chains," *Electronic Journal of Probability*, vol. 13, no. 34, pp. 1000–1034, 2008.

[2] C. Andrieu, A. Doucet, and R. Holenstein, "Particle Markov chain Monte Carlo methods," *Journal Royal Statistical Society B*, vol. 72, no. 3, pp. 269–342, 2010.

[3] N. Bellomo and M. Pulvirenti, "Generalized kinetic models in applied sciences," in *Modeling in Applied Sciences,* Modeling and Simulation in Science, Engineering, and Technology, pp. 1–19, Birkhäuser, Boston, 2000.

[4] N. Bellomo and M. Pulvirenti, eds., *Modeling in Applied Sciences,* Modeling and Simulation in Science, Engineering, and Technology. Birkhäuser, Boston, 2000.

[5] E. Cancès, B. Jourdain, and T. Lelièvre, "Quantum Monte Carlo simulations of fermions. A mathematical analysis of the fixed-node approximation," *ESAIM: M2AN*, vol. 16, no. 9, pp. 1403–1449, 2006.

[6] R. Carmona, J.-P. Fouque, and D. Vestal, "Interacting particle systems for the computation of rare credit portfolio Losses," *Finance and Stochastics*, vol. 13, no. 4, pp. 613–633, 2009.

[7] F. Caron, P. Del Moral, A. Tantar, and E. Tantar, "Simulation particulaire, Dimensionnement en conditions extrêmes," Research contract no. 2010-IFREMER-01 with IFREMER, (92p) September, 2010.

[8] F. Cerou, P. Del Moral, Fr. Le Gland, and P. Lezaud, "Genealogical models in entrance times rare event analysis," *Alea, Latin American Journal of Probability And Mathematical Statistics*, vol. 1, pp. 181–203, 2006.

[9] F. Cerou, P. Del Moral, T. Furon, and A. Guyader, "Sequential Monte Carlo for rare event estimation," *Statistics and Computing*, to appear, 2011.

[10] F. Cerou, P. Del Moral, A. Guyader, F. LeGland, P. Lezaud, and H. Topart, "Some recent improvements to importance splitting," in (preliminary version), *Proceedings of 6th International Workshop on Rare Event Simulation*, Bamberg, Germany, 2006.

[11] D. A. Dawson, "Stochastic evolution equations and related measure processes," *J. Multivariate Anal.*, vol. 3, pp. 1–52, 1975.

[12] D. A. Dawson, "Measure-valued Markov processes," in *École d'Été de Probabilités de Saint-Flour 1991,* Lecture Notes Math. 1541, pp. 1–260, Springer, Berlin, 1993.

[13] D. A. Dawson and P. Del Moral, "Large deviations for interacting processes in the strong topology," in *Statistical Modeling and Analysis for Complex Data Problem*, (P. Duchesne and B. Rmillard, eds.), pp. 179–209, Springer, 2005.

[14] D. A. Dawson, K. J. Hochberg, and V. Vinogradov, "On weak convergence of branching particle systems undergoing spatial motion," in *Stochastic Analysis: Random Fields and Measure-valued Processes*, pp. 65–79, *Israel Math. Conf. Proc. 10. Bar-Ilan Univ., Ramat Gan.*, 1996.

[15] P. Del Moral, "Non linear filtering: Interacting particle solution," *Markov Processes and Related Fields*, vol. 2, no. 4, pp. 555–580, 1996.

[16] P. Del Moral, "Measure valued processes and interacting particle systems. Application to nonlinear filtering problems," *Annals of Applied Probability*, vol. 8, no. 2, pp. 438–495, 1998.

[17] P. Del Moral, *Feynman-Kac Formulae: Genealogical and Interacting Particle Systems with Applications*. New York: Springer-Verlag, 2004.

[18] P. Del Moral and A. Doucet, "On a class of genealogical and interacting Metropolis Models," in *Séminaire de Probabilités* XXXVII, (J. Azéma, M. Emery, M. Ledoux, and M. Yor, eds.), pp. 415–446, Lecture Notes in Mathematics 1832, Springer-Verlag, Berlin, 2003.

[19] P. Del Moral and A. Doucet, "Particle motions in absorbing medium with hard and soft obstacles," *Stochastic Analysis and Applications*, vol. 22, no. 5, pp. 1175–1207, 2004.

[20] P. Del Moral and A. Doucet, *Sequential Monte Carlo & Genetic Particle Models. Theory and Practice*. Chapman & Hall, Green series, Statistics and Applied Probability, to appear, 2012.

[21] P. Del Moral, A. Doucet, and A. Jasra, "Sequential Monte Carlo samplers," *J. Royal Statist. Soc. B*, vol. 68, pp. 411–436, 2006.

[22] P. Del Moral, A. Doucet, and A. Jasra, "An adaptive sequential Monte Carlo method for approximate Bayesian computation," *Research report, Imperial College*, 2008.

[23] P. Del Moral, A. Doucet, and S. S. Singh, "A backward particle interpretation of Feynman-Kac formulae," *ESAIM M2AN*, vol. 44, no. 5, pp. 947–976, 2010.

[24] P. Del Moral, A. Doucet, and S. S. Singh, "Forward smoothing using sequential Monte Carlo," Technical Report 638. Cambridge University Engineering Department., 2010.

[25] P. Del Moral, A. Doucet, and S. S. Singh, "Uniform stability of a particle approximation of the optimal filter derivative," Technical Report CUED/F-INFENG/TR 668. Cambridge University Engineering Department., 2011.

[26] P. Del Moral and J. Garnier, "Genealogical particle analysis of rare events," *Annals of Applied Probability*, vol. 15, no. 4, pp. 2496–2534, 2005.
[27] P. Del Moral and J. Garnier, "Simulations of rare events in fiber optics by interacting particle systems," (preliminary version), *Optics Communications*, Elsevier, 2006.
[28] P. Del Moral and A. Guionnet, "Large deviations for interacting particle systems. Applications to nonlinear filtering problems," *Stochastic Processes Appl.*, vol. 78, pp. 69–95, 1998.
[29] P. Del Moral and A. Guionnet, "A central limit theorem for nonlinear filtering using interacting particle systems," *Ann. Appl. Probab.*, vol. 9, no. 2, pp. 275–297, 1999.
[30] P. Del Moral and A. Guionnet, "On the stability of measure valued processes with applications to filtering," *C.R. Acad. Sci. Paris*, vol. t. 329, no. Serie I, pp. 429–434, 1999.
[31] P. Del Moral and A. Guionnet, "On the stability of interacting processes with applications to filtering and genetic algorithms," *Ann. Inst. Henri Poincaré*, vol. 37, no. 2, pp. 155–194, 2001.
[32] P. Del Moral, S. Hu, and L. M. Wu, "Moderate deviations for mean field particle models," in preparation, 2011.
[33] P. Del Moral and J. Jacod, "The Monte-Carlo method for filtering with discrete-time observations: Central limit theorems," in *Numerical Methods and Stochastics (Toronto, ON, 1999)*, Vol. 34 of Fields Inst. Commun., pp. 29–53, American Mathematical Society, Providence, RI, 2002.
[34] P. Del Moral, J. Jacod, and P. Protter, "The Monte-Carlo method for filtering with discrete-time observations," *Probability Theory and Related Fields*, vol. 120, pp. 346–368, 2001.
[35] P. Del Moral, M. A. Kouritzin, and L. Miclo, "On a class of discrete generation interacting particle systems," *Electronic Journal of Probability*, vol. 16, pp. 1–26, 2001.
[36] P. Del Moral and M. Ledoux, "On the convergence and the applications of empirical processes for interacting particle systems and nonlinear filtering," *Journal of Theoretical Probability*, vol. 13, no. 1, pp. 225–257, 2000.
[37] P. Del Moral and P. Lezaud, "Branching and interacting particle interpretation of rare event probabilities," in *Stochastic Hybrid Systems: Theory and Safety Critical Applications*, (H. Blom and J. Lygeros, eds.), Springer-Verlag, Heidelberg, 2006.
[38] P. Del Moral and L. Miclo, "Branching and interacting particle systems approximations of Feynman-Kac formulae with applications to nonlinear filtering," in *Seminaire de Probabilités* XXXIV, Vol. 1729, (J. Azma, M. Emery, M. Ledoux, and M. Yor, eds.), pp. 1–145, Lecture Notes in Mathematics, Springer-Verlag Berlin, 2000.
[39] P. Del Moral and L. Miclo, "Particle approximations of Lyapunov exponents connected to Schrödinger operators and Feynman-Kac semigroups," *ESAIM Probability and Statistics*, vol. 7, pp. 169–207, 2003.
[40] P. Del Moral and E. Rio, "Concentration inequalities for mean field particle models," *Annals of Applied Probability*, vol. 21, no. 3, pp. 1017–1052, 2011.

[41] P. Del Moral and T. Zajic, "A note on Laplace-Varadhan's integral lemma," *Bernoulli*, vol. 9, no. 1, pp. 49–65, 2003.
[42] G. B. Di Masi, M. Pratelli, and W. G. Runggaldier, "An approximation for the nonlinear filtering problem with error bounds," *Stochastics*, vol. 14, no. 4, pp. 247–271, 1985.
[43] R. Douc, A. Guillin, and J. Najim, "Moderate deviations for particle filtering," *The Annals of Applied Probability*, vol. 15, no. 1B, pp. 587–614, 2005.
[44] A. Doucet, N. De Freitas, and N. Gordon, eds., *Sequential Monte Carlo Methods in Pratice,* Statistics for Engineering and Information Science. New York: Springer, 2001.
[45] E. B. Dynkin, "An introduction to branching measure-valued processes," in *CRM Monograph Series*, Vol. 6. Amer. Math. Soc., Providence, 1994.
[46] M. El Makrini, B. Jourdain, and T. Lelièvre, "Diffusion Monte Carlo method: Numerical analysis in a simple case," *ESAIM: M2AN*, vol. 41, no. 2, pp. 189–213, 2007.
[47] S. N. Ethier and T. G. Kurtz, *Markov Processes: Characterization and Convergence*. Wiley, New York, 1995.
[48] C. J. Everett and S. Ulam, *Proc. Nat. Acad. Sciences*, vol. 403, no. 33, 1948.
[49] C. J. Everett and S. Ulam, *USAEC, Los Alamos report, LADC-533, LADC-534*. declassified, 1948.
[50] W. Feller, "Diffusion processes in genetics," in *Proc. Second Berkeley Symp. Math. Statist. Prob.*, pp. 227–246, Berkeley: University of California Press, 1951.
[51] P. A. Ferrari, H. Kesten, S. Martínez, and P. Picco, "Existence of quasi-stationary distributions. A renewal dynamical approach," *Ann. Probab.*, vol. 23, pp. 501–521, 1995.
[52] Fr. Caron, P. Del Moral, A. Doucet, and M. Pace. On the Conditional Distributions of Spatial Point Processes *HAL-INRIA RR-7232*. To appear in *Advances in Applied Probability*, 2011.
[53] Fr. Caron, P. Del Moral, A. Doucet, and M. Pace, "Particle approximations of a class of branching distribution flows arising in multi-target tracking," *HAL-INRIA RR-7233 (2010) [29p], SIAM Journal on Control and Optimization*, vol. 49, no. 4, pp. 1766–1792, 2011.
[54] Fr. Caron, P. Del Moral, M. Pace, and B. N. Vo, "On the stability and the approximation of branching distribution flows, with applications to nonlinear multiple target filtering," *Stoch. Analysis and Applications*, vol. 29, no. 6, pp. 951–997, 2011.
[55] Fr. Cerou, A. Guyader, R. Rubinstein, and R. Vaismana, "On the use of smoothing to improve the performance of the splitting method," *Stochastic Models*, vol. 27, no. 4, pp. 629–650, 2011.
[56] E. Giné, "Lectures on some aspects of the bootstrap," in *Lecture Notes in Maths*, Vol. 1665, Springer, 1997.
[57] F. Gosselin, "Asymptotic behavior of absorbing Markov chains conditional on nonabsorption for applications in conservation biology," *Ann. Appl. Probab.*, vol. 11, pp. 261–284, 2001.
[58] C. Graham and S. Méléard, "Stochastic particle approximations for generalized Boltzmann models and convergence estimates," *Ann. Probab.*, vol. 25, no. 1, pp. 115–132, 1997.

[59] C. Graham and S. Méléard, "Probabilistic tools and Monte-Carlo approximations for some Boltzmann equations," in *CEMRACS 1999 (Orsay)*, Vol. 10 of ESAIM Proceedings, pp. 77–126, (electronic). Société de Mathématiques Appliquées et Industrielles, Paris, 1999.

[60] A. Guyader, N. Hengartner, and E. Matzner-Løber, "Simulation and estimation of extreme quantiles and extreme probabilities," *Applied Mathematics & Optimization*, vol. 64, no. 2, pp. 171–196, 2011.

[61] H. P. McKean, Jr, "A class of Markov processes associated with nonlinear parabolic equations," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 56, pp. 1907–1911, 1966.

[62] J. H. Holland, *Adaptation in Natural and Artificial Systems*. University of Michigan Press, Ann Arbor, 1975.

[63] H.P. McKean, Jr, "Propagation of chaos for a class of non-linear parabolic equations," in *Stochastic Differential Equations (Lecture Series in Differential Equations, Session 7, Catholic University, 1967)*, pp. 41–57, Air Force Office of Scientific Research, Arlington, VA, 1967.

[64] A. M. Johansen, P. Del Moral, and A. Doucet, "Sequential Monte Carlo samplers for rare events," in (preliminary version), *Proceedings of 6th International Workshop on Rare Event Simulation*, Bamberg, Germany, 2006.

[65] W. B. Johnson, G. Schechtman, and J. Zinn, "Best constants in moment inequalities for linear combinations of independent and exchangeable random variables," *Ann. Probab.*, vol. 13, no. 1, pp. 234–253, 1985.

[66] M. Kac, "On distributions of certain Wiener functionals," *Trans. American Math. Soc.*, vol. 61, no. 1, pp. 1–13, 1949.

[67] T. Klein and E. Rio, "Concentration around the mean for maxima of empirical processes," *Ann. Probab.*, vol. 33, no. 3, pp. 1060–1077, 2005.

[68] H. Korezlioglu and W. J. Runggaldier, "Filtering for nonlinear systems driven by nonwhite noises: An approximating scheme," *Stochastics and Stochastics Rep.*, vol. 44, no. 1–2, pp. 65–102, 1983.

[69] J. F. Le Gall, "Spatial branching processes, random snakes, and partial different equations," in *Lectures Given in the Nachdiplomsvorlesung, ETH Zürich*, Published in the series, Lectures in Mathematics, ETH Zürich, Birkhauser, 1999.

[70] M. Ledoux, "On Talagrand's deviation inequalities for product measures," *ESAIM Probability and Statistics*, vol. 1, pp. 63–87, 1996.

[71] M. Ledoux, "The concentration of measure phenomenon," AMS Monographs, Providence, 2001.

[72] M. Ledoux and M. Talagrand, *Probability in Banach Spaces*. Springer-Verlag, 1991.

[73] R. Mahler, "A theoretical foundation for the Stein-Winter Probability Hypothesis Density (PHD) multi-target tracking approach," in *Proc. MSS Nat'l Symp. on Sensor and Data Fusion*, Vol. 1, 2000.

[74] R. Mahler, "Multi-target Bayes filtering via first-order multi-target moments," *IEEE Trans. Aerospace & Electronic Systems*, vol. 39, no. 4, pp. 1152–1178, 2003.

[75] S. Méléard, "Asymptotic behaviour of some interacting particle systems; McKean-Vlasov and Boltzmann models," in *Probabilistic Models for Nonlinear*

Partial Differential Equations, Montecatini Terme, 1995, (D. Talay and L. Tubaro, eds.), Lecture Notes in Mathematics 1627. Springer-Verlag, Berlin, 1996.

[76] S. Méléard, "Convergence of the fluctuations for interacting diffusions with jumps associated with Boltzmann equations," *Stochastics Stochastics Rep.*, vol. 63, no. 3–4, pp. 195–225, 1998.

[77] S. Méléard, "Probabilistic interpretation and approximations of some Boltzmann equations," in *Stochastic models (Spanish) (Guanajuato, 1998)*, Vol. 14 of *Aportaciones Mat. Investig.*, pp. 1–64, Soc. Mat. Mexicana, México, 1998.

[78] S. Méléard, "Stochastic approximations of the solution of a full Boltzmann equation with small initial data," *ESAIM Probab. Stat.*, vol. 2, pp. 23–40, 1998.

[79] N. Metropolis and S. Ulam, "The Monte Carlo method," *Journal of the American Statistical Association*, vol. 44, no. 247, pp. 335–341, 1949.

[80] J. Picard, "Approximation of the non linear filtering problems and order of convergence," in *Lecture Notes in Control and Inf. Sc*, Vol. 61, 1984.

[81] D. Pollard, *Empirical Processes: Theory and Applications,* Vol. 2. NSF-CBMS Regional Conference Series in Probability and Statistics. Publication of the Institute of Mathematical Statistics. Hayward, CA, 1990.

[82] M. Pulvirenti, "Kinetic limits for stochastic particle systems," in *Probabilistic Models for Nonlinear Partial Differential Equations (Montecatini Terme, 1995)*, Vol. 1627 of Lecture Notes in Mathematics, pp. 96–126, Springer, Berlin, 1996.

[83] S. T. Rachev, *Probability Metrics and the Stability of Stochastic Models.* Wiley, New York, 1991.

[84] E. Rio, "Local invariance principles and their applications to density estimation," *Probability and related Fields*, vol. 98, pp. 21–45, 1994.

[85] M. Rousset, "On the control of an interacting particle approximation of Schrödinger ground states," *SIAM J. Math. Anal.*, vol. 38, no. 3, pp. 824–844, 2006.

[86] W. Runggaldier and L. Stettner, *Approximations of Discrete Time Partially Observed Control Problems.* Applied Mathematics Monographs CNR, Giardini Editori, Pisa, 1994.

[87] Z. Skolicki and K. De Jong, "The influence of migration sizes and intervals on island models," in *Proceedings of the Genetic and Evolutionary Computation Conference (GECCO-2005)*, ACM Press, 2005.

[88] D. Steinsaltz and S. N. Evans, "Markov mortality models: Implications of quasistationarity and varying initial distributions," *Theor. Pop. Biol.*, vol. 65, pp. 319–337, 2004.

[89] A. S. Sznitman, "Topics in propagation of chaos," in *Ecole d'Eté de Probabilités de Saint-Flour XIX-1989*, (P. L. Hennequin, ed.), Lecture Notes in Mathematics 1464. Springer-Verlag, Berlin, 1991.

[90] M. Talagrand, "Concentration of measure and isoperimetric inequalities in product spaces," *Inst. Hautes Études Sci. Publ. Math.*, vol. 81, pp. 73–205, 1995.

[91] M. Talagrand, "New concentration inequalities in product spaces," *Invent. Math.*, vol. 125, pp. 505–563, 1996.

[92] M. Talagrand, "A new look at independence," *Ann. Probab.*, vol. 24, pp. 1–34, 1996.

[93] A. N. Van der Vaart and J. A. Wellner, *Weak Convergence and Empirical Processes with Applications to Statistics*. New York: Springer Series in Statistics. Springer, 1996.

[94] J. Vanneste, "Estimating generalized Lyapunov exponents for products of random matrices," *Phys. Rev. E*, vol. 81, no. 3, p. 036701, 2010.

[95] J.-P. Vila and V. Rossi, "Nonlinear filtering in discrete time: A particle convolution approach, Biostatistic group of Montpellier," Technical Report 04-03, (available at http://vrossi.free.fr/recherche.html), 2004.

[96] D. Whitley, S. Rana, and R. B. Heckendorn, "The island model genetic algorithm: On separability, population size and convergence," *Journal of Computing and Information Technology*, vol. 7, no. 1, pp. 33–47, 1999.