Foundations and Trends® in
Machine Learning
3:3-4

On the Concentration Properties of Interacting Particle Processes

Pierre Del Moral, Peng Hu and Liming Wu

now
the essence of knowledge

On the Concentration Properties of Interacting Particle Processes

Foundations and Trends® in Machine Learning

Published, sold and distributed by:
now Publishers Inc.
PO Box 1024
Hanover, MA 02339
USA
Tel. +1-781-985-4510
www.nowpublishers.com
sales@nowpublishers.com

Outside North America:
now Publishers Inc.
PO Box 179
2600 AD Delft
The Netherlands
Tel. +31-6-51115274

The preferred citation for this publication is P. Del Moral, P. Hu, and L. Wu, On the Concentration Properties of Interacting Particle Processes, Foundation and Trends® in Machine Learning, vol 3, nos 3–4, pp 225–389, 2010.

ISBN: 978-1-60198-512-5

On the Concentration Properties of Interacting Particle Processes

Pierre Del Moral

INRIA, Université Bordeaux 1
France

Peng Hu

INRIA, Université Bordeaux 1
France

Liming Wu

Institute of Applied Mathematics
AMSS, Chinese Academy of Sciences
China

Université Blaise Pascal
France

now
the essence of knowledge

Boston – Delft

Foundations and Trends® in Machine Learning

Volume 3 Issues 3–4, 2010

Editorial Scope

Foundations and Trends® in Machine Learning will publish survey and tutorial articles in the following topics:

- Adaptive control and signal processing
- Applications and case studies
- Behavioral, cognitive and neural learning
- Bayesian learning
- Classification and prediction
- Clustering
- Data mining
- Dimensionality reduction
- Evaluation
- Game theoretic learning
- Graphical models
- Independent component analysis
- Inductive logic programming
- Kernel methods
- Markov chain Monte Carlo
- Model choice
- Nonparametric methods
- Online learning
- Optimization
- Reinforcement learning
- Relational learning
- Robustness
- Spectral methods
- Statistical learning theory
- Variational inference
- Visualization

Information for Librarians

Foundations and Trends® in Machine Learning, 2010, Volume 3, 4 issues. ISSN paper version 1935-8237. ISSN online version 1935-8245. Also available as a combined paper and online subscription.

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

DOI: 10.1561/2200000026

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.

Contents

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 ∞.