



IN PARTNERSHIP WITH:
CNRS

Université de Bordeaux

Activity Report 2011

Project-Team ALEA

Advanced Learning Evolutionary Algorithms

IN COLLABORATION WITH: Institut de Mathématiques de Bordeaux (IMB)

RESEARCH CENTER
Bordeaux - Sud-Ouest

THEME
Stochastic Methods and Models

Table of contents

1. Members	1
2. Overall Objectives	1
2.1. Overall Objectives	1
2.2. Highlights	2
3. Scientific Foundations	2
4. Application Domains	3
5. New Results	4
5.1. Snell envelope with small probability criteria	4
5.2. On the concentration properties of Interacting particle processes	5
5.3. A Robbins-Monro procedure for estimation in semiparametric regression models	5
5.4. Discrete Time Markovian Agents Interacting Through a Potential	5
5.5. An Adaptive Interacting Wang-Landau Algorithm for Automatic Density Exploration	5
5.6. A sharp analysis on the asymptotic behavior of the Durbin-Watson statistic for the first-order autoregressive process	6
5.7. Large deviations for Gaussian stationary processes and semi-classical analysis	6
5.8. Sharp large deviations for the non-stationary Ornstein-Uhlenbeck process	6
5.9. Markovian quadratic and superquadratic BSDEs with an unbounded terminal condition	6
5.10. New approach on recursive and non-recursive SIR	7
5.11. Classification of EEG data by evolutionary algorithm for the study of vigilance states	7
5.12. Genetic Programming	7
5.13. Relevance of the Holderian regularity-based interpolation for range-Doppler ISAR image post-processing	8
5.14. On-line changepoint detection and parameter estimation with application to genomic data	8
5.15. Bayesian Sparsity-Path-Analysis of Genetic Association Signal using Generalized t Priors	8
6. Contracts and Grants with Industry	9
6.1. Contract with EDF	9
6.2. Contract with EDF	9
6.3. Contract with CEA	9
7. Partnerships and Cooperations	9
7.1. Regional Initiatives	9
7.2. National Initiatives	9
7.2.1. ANR Propagation (2010-2012)	9
7.2.2. Project PEP2	10
7.3. European Initiatives	10
7.4. International Initiatives	10
8. Dissemination	10
8.1. Animation of the scientific community	10
8.2. Teaching	11
9. Bibliography	12

Project-Team ALEA

Keywords: Statistical Learning, Particle Filtering, Rare Events, Stochastic Algorithms, Monte Carlo Methods, Evolutionary Algorithms

1. Members

Research Scientists

Pierre Del Moral [INRIA, Team leader, Senior Researcher (DR), HdR]
François Caron [INRIA, Junior Researcher (CR)]

Faculty Members

Bernard Berçu [University Bordeaux 1, Professor (PR), HdR]
Pierrick Legrand [University Bordeaux 2, Associate Professor (MdC)]
Adrien Richou [University Bordeaux 1, Associate Professor (MdC), since Sept. 2011]

External Collaborators

Arnaud Doucet [University of Oxford]
Josselin Garnier [University Paris 7]
Pierre Minvielle [CEA]

Technical Staff

Adrien Todeschini [INRIA, funding from INRIA ADT BiiPS]
Denis Arrivault [INRIA, funding from ANR Propagation]
Jérémie Houssineau [INRIA, funding from ANR Propagation, April-May 2011]

PhD Students

Michele Pace [INRIA, funding from INRIA, Sept. 2008 - August 2011]
Aurélié Le Cain [CEA, CIFRE, since Sept. 2008]
François Giraud [CEA, CIFRE, since Nov. 2009]
Hu Peng [University of Bordeaux, funding from French Research Ministry, since Oct. 2009]
Philippe Fraysse [University of Bordeaux, funding from ENS Cachan, since Sept. 2010]
Paul Lemaître [EDF, CIFRE, since Sept. 2010]
Frédéric Proia [INRIA, funding from Région Aquitaine and INRIA, since Dec. 2010]
Nicolas Antunes [University of Bordeaux, funding from French Research Ministry, since Sept. 2011]
Vassili Blandin [University of Bordeaux, funding from ENS Cachan, since Sept. 2011]

Administrative Assistant

Nicolas Jahier [Team Assistant]

2. Overall Objectives

2.1. Overall Objectives

In recent years, a new generation of numerical algorithms have begun to spread through the scientific community. Surprisingly enough, up to a few exceptions, many of these modern ideas do not really come from physics, but from biology and ethology. In a growing number of scientific disciplines, the researchers are now interpreting real world processes and engineering type systems less like purely deterministic and crude clockwork mechanisms, but much more like random and sophisticated biology inspired processes. This new generation of engineering models is based on stochastic ideas and natural principles like : chance, randomness, interactions, reinforcement strategies, exploration rules, biology-inspired adaptation and selection transitions, learning, reproduction, birth and death, ancestral lines and historical processes, genealogical tree evolutions, as well as self-organization principles, and many others.

These biology-inspired stochastic algorithms are often presented as natural heuristic simulation schemes without any mathematical foundations, nor a single performance analysis ensuring their convergence, nor even a single theoretical discussion that clarifies the applicability of these models. An important aspect of our project is to create a concrete bridge between pure and applied probability, statistics, biology, stochastic engineering and computer sciences. This fundamental bridging effort is probably one of the most important key to turn real nature's processes into engineering devices and stochastic algorithms, by learning what can be abstracted, copied or adapted. In the reverse angle, we can mention that these abstracted models adapting nature mechanisms and biological capabilities also provides a better understanding of the real processes.

By essence, *the team-project is not a single application-driven research project*. The reasons are three-fold. Firstly, the same stochastic algorithm is very often used in a variety of application areas. On the other hand every application domain area offers a series of different perspectives that can be used to improve the design and the performances of the algorithms. Last but not least, concrete industrial applications, as well as most of concrete problems arising in biology, physics and chemistry, require a specific attention. In general, we do not use a single class of stochastic algorithm but a broader set of stochastic search algorithms that incorporates facets of nature inspired strategies.

Our research project is centered on two central problems in advanced stochastic engineering: *Bayesian inference and rare event simulation* and more particularly *unsupervised learning, multi-target tracking, data assimilation and forecasting, as well as infection spreads inference*. These important and natural research directions have emerged as logical parts of the team project combined with interdisciplinary approaches well-represented at Bordeaux university campus.

The fundamental and the theoretical aspects of our research project are essentially concerned with the stochastic analysis of the following three classes of biology inspired stochastic algorithms: *branching and interacting particle systems, reinforced random walks and self-interacting processes, random tree based models*. One of our prospective research project is to apply the Bayesian learning methodology and the recent particle filter technology to the design of a new generation of interactive **evolutionary computation** and **stochastic art composition** models.

2.2. Highlights

The team co-organized the **Machine Learning Summer School** in Bordeaux, from September 4 to September 17 2011. This event gathered 100 participants, mostly PhD student, from 18 different countries and covered various topics (support vector machines, Monte Carlo methods, Bayesian inference, boosting, etc.) presented by world-class experts in their field. MLSS is a recurrent and important event of the machine learning community.

3. Scientific Foundations

3.1. Fondations

This idea of analyzing nature systems and transferring the underlying principles into stochastic algorithms and technical implementations is one of the central component of the ALEA team project. Adapting nature mechanisms and biological capabilities clearly provides a better understanding of the real processes, and it also improves the performance and the power of engineers devices. Our project is centered on both the understanding of biological processes in terms of mathematical, physical and chemical models, and on the other hand, on the use of these biology inspired stochastic algorithms to solve complex engineering problems.

There is a huge series of virtual interfaces, robotic devices, numerical schemes and stochastic algorithms which were invented mimicking biological processes or simulating natural mechanisms. The terminology "*mimicking or simulating*" doesn't really mean to find an exact copy of natural processes, but *to elaborate the mathematical principles so that they can be abstracted from the original biological or physical model*. In our context, the whole series of evolutionary type principles discussed in previous sections can be abstracted into only three different and natural classes of stochastic algorithms, depending on the nature of the biology-inspired interaction mechanism used in the stochastic evolution model. These three stochastic search models are listed below :

1) *Branching and interacting particle systems (birth and death chains, spatial branching processes, mean-field interaction between generations):*

The first generation of adaptive branching-selection algorithms is very often built on the same genetic type paradigm: When exploring a state space with many particles, we duplicate better fitted individuals at the expense of light particles with poor fitness die. From a computational point of view, we generate a large number of random problem solvers. Each one is then rated according to a fitness or performance function defined by the developer. Mimicking natural selection, an evolutionary algorithm selects the best solvers in each generation and breeds them.

2) *Reinforced random walks and self-interacting chains (reinforced learning strategies, interaction processes with respect to the occupation measure of the past visited sites):*

This type of reinforcement is observed frequently in nature and society, where "beneficial" interactions with the past history tend to be repeated. A new class of historical mean field type interpretation models of reinforced processes were developed by the team project leader in a pair of articles [48], [47]. Self interaction gives the opportunity to build new stochastic search algorithms with the ability to, in a sense, re-initialized their exploration from the past, re-starting from some better fitted initial value already met in the past [49], [50].

3) *Random tree based stochastic exploration models (coalescent and genealogical tree search explorations techniques on path space):*

The last generation of stochastic random tree models is concerned with biology-inspired algorithms on paths and excursions spaces. These genealogical adaptive search algorithms coincide with genetic type particle models in excursion spaces. They have been applied with success in generating the excursion distributions of Markov processes evolving in critical and rare event regimes, as well as in path estimation and related smoothing problems arising in advanced signal processing (cf. [45] and references therein). We underline the fact that the complete mathematical analysis of these random tree models, including their long time behavior, their propagations of chaos properties, as well as their combinatorial structures are far from being completed. This class of genealogical tree based models has been introduced in [46] for solving smoothing problems and more generally Feynman-Kac semigroups on path spaces, see also [44], [45], and references therein.

4. Application Domains

4.1. Application Domains

This short section is only concerned with the list of concrete application domains developed by our team project on Bayesian inference and unsupervised learning, nonlinear filtering and rare event analysis. Most of these application areas result from fruitful collaborations with other national institutes through a series of four recently selected ANR research projects and one INRIA-INRA joint research project.

Three application domains are directly related to evolutionary computing, particle filtering and Bayesian inference. They are currently investigated by our team project:

1. **Multi-target tracking.** Multi-target tracking deals with the task of estimating the states of a set of moving targets from a set of measurements obtained sequentially. These measurements may either arise from one of the targets or from clutter and the measurement-to-target association is generally

unknown. This problem can then be recast as a dynamic clustering one where the clusters are the clutter and the different targets. The targets actually move in time, some targets may appear/disappear over time and the number of targets is generally unknown and time-varying. We are running this research project with the DCNS-SIS division in Toulon.

2. **Forecasting and Data assimilation :** This new application domain concerns the application of the particle filter technology and more general sequential Monte Carlo methods to data assimilation problems arising in forecasting. The ALEA team project is involved in the [ANR 2008 selected project PREVASSEMBLE](#) with Météo France Toulouse, the INRIA Rennes and the LMD in Paris.
3. **Virtual prairie:** This application domain of evolutionary computing is concerned with the design of ecological systems, mixed-species models and prairial ecosystems. For more details, we refer the reader to the web site of the [Virtual Prairie project](#) The ALEA project is a partner of the 2008 ANR SYSCOM project named MODECOL.

Three other application domains are directly related to rare event analysis using particle stochastic simulations techniques. These projects are currently investigated by our team project, two of them are 2008's ANR research projects :

1. **Watermarking of digital contents:**

The terminology watermarking refers to a set of techniques for imbedding/hiding information in a digital audio or video file, such that the change is not noticed, and very hard to remove. In order to be used in an application, a watermarking technique must be reliable. Protection false alarms and failures of traceability codes are practically not achievable without using rare event analysis. This application domain area of particle rare event technology is the subject of joint ANR 2007 research project with the IRISA-INRIA in Rennes and the LIS INPG in Grenoble.

2. **Epidemic propagations analysis:**

This project aims at developing stochastic mathematical models for the spread of transmissible infectious diseases, together with dedicated statistical methodologies, with the intent to deliver efficient diagnostic/prediction tools for epidemiologists. This application domain area of particle rare event technology is the subject of joint ANR 2008 research project with Telecom Paristech, the Laboratoire Paul Painlevé in Lille 1 and the University of Paris 5 (cf. [Programme Systèmes Complexes et Modélisation Mathématique, list of 2008 selected projects](#)).

3. **Statistical eco-microbiology predictions:**

This project aims at developing stochastic models and algorithms for the analysis of bacteriology ecosystems, especially in food safety. The objective is to predict and control critical risk of proliferations. This is the subject of joint research project with the INRA of Paris and Montpellier (Appel d'Offre INRIA-INRA 2008 : Systèmes Complexes).

5. New Results

5.1. Snell envelope with small probability criteria

We present in [33] a new algorithm to compute the Snell envelope in the specific case where the criteria to optimize is associated with a small probability or a rare event. This new approach combines the Stochastic Mesh approach of Broadie and Glasserman with a particle approximation scheme based on a specific change of measure designed to concentrate the computational effort in regions pointed out by the criteria. The theoretical analysis of this new algorithm provides non asymptotic convergence estimates. Finally, the numerical tests confirm the practical interest of this approach.

5.2. On the concentration properties of Interacting particle processes

In [34], we present 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 the quantitative performance analysis of a series of advanced Monte Carlo methods, including particle filters, genetic type island models, Markov bridge models, interacting particle Markov chain Monte Carlo methodologies.

5.3. A Robbins-Monro procedure for estimation in semiparametric regression models

The paper [39] is devoted to the parametric estimation of a shift together with the nonparametric estimation of a regression function in a semiparametric regression model. We implement a Robbins-Monro procedure very efficient and easy to handle. On the one hand, we propose a stochastic algorithm similar to that of Robbins-Monro in order to estimate the shift parameter. A preliminary evaluation of the regression function is not necessary for estimating the shift parameter. On the other hand, we make use of a recursive Nadaraya-Watson estimator for the estimation of the regression function. This kernel estimator takes in account the previous estimation of the shift parameter. We establish the almost sure convergence for both Robbins-Monro and Nadaraya-Watson estimators. The asymptotic normality of our estimates is also provided.

5.4. Discrete Time Markovian Agents Interacting Through a Potential

A discrete time stochastic model for a multiagent system given in terms of a large collection of interacting Markov chains is studied. The evolution of the interacting particles is described through a time inhomogeneous transition probability kernel that depends on the 'gradient' of the potential field. The particles, in turn, dynamically modify the potential field through their cumulative input. Interacting Markov processes of the above form have been suggested as models for active biological transport in response to external stimulus such as a chemical gradient. One of the basic mathematical challenges is to develop a general theory of stability for such interacting Markovian systems and for the corresponding nonlinear Markov processes that arise in the large agent limit. Such a theory would be key to a mathematical understanding of the interactive structure formation that results from the complex feedback between the agents and the potential field. It will also be a crucial ingredient in developing simulation schemes that are faithful to the underlying model over long periods of time. The goal of the work developed in [42] is to study qualitative properties of the above stochastic system as the number of particles (N) and the time parameter (n) approach infinity. In this regard asymptotic properties of a deterministic nonlinear dynamical system, that arises in the propagation of chaos limit of the stochastic model, play a key role. We show that under suitable conditions this dynamical system has a unique fixed point. This result allows us to study stability properties of the underlying stochastic model. We show that as $N \rightarrow \infty$, the stochastic system is well approximated by the dynamical system, uniformly over time. As a consequence, for an arbitrarily initialized system, as $N \rightarrow \infty$ and $n \rightarrow \infty$, the potential field and the empirical measure of the interacting particles are shown to converge to the unique fixed point of the dynamical system. In general, simulation of such interacting Markovian systems is a computationally daunting task. We propose a particle based approximation for the dynamic potential field which allows for a numerically tractable simulation scheme. It is shown that this simulation scheme well approximates the true physical system, uniformly over an infinite time horizon.

5.5. An Adaptive Interacting Wang-Landau Algorithm for Automatic Density Exploration

While statisticians are well-accustomed to performing exploratory analysis in the modeling stage of an analysis, the notion of conducting preliminary general-purpose exploratory analysis in the Monte Carlo

stage (or more generally, the model-fitting stage) of an analysis is an area which we feel deserves much further attention. Towards this aim, the paper [41] proposes a general-purpose algorithm for automatic density exploration. The proposed exploration algorithm combines and expands upon components from various adaptive Markov chain Monte Carlo methods, with the Wang-Landau algorithm at its heart. Additionally, the algorithm is run on interacting parallel chains – a feature which both decreases computational cost as well as stabilizes the algorithm, improving its ability to explore the density. Performance is studied in several applications. Through a Bayesian variable selection example, the authors demonstrate the convergence gains obtained with interacting chains. The ability of the algorithm’s adaptive proposal to induce mode-jumping is illustrated through a trimodal density and a Bayesian mixture modeling application. Lastly, through a 2D Ising model, the authors demonstrate the ability of the algorithm to overcome the high correlations encountered in spatial models.

5.6. A sharp analysis on the asymptotic behavior of the Durbin-Watson statistic for the first-order autoregressive process

The purpose of the paper [40] is to provide a sharp analysis on the asymptotic behavior of the Durbin-Watson statistic. We focus our attention on the first-order autoregressive process where the driven noise is also given by a first-order autoregressive process. We establish the almost sure convergence and the asymptotic normality for both the least squares estimator of the unknown parameter of the autoregressive process as well as for the serial correlation estimator associated to the driven noise. In addition, the almost sure rates of convergence of our estimates are also provided. It allows us to establish the almost sure convergence and the asymptotic normality for the Durbin-Watson statistic. Finally, we propose a new bilateral statistical test for residual autocorrelation.

5.7. Large deviations for Gaussian stationary processes and semi-classical analysis

In [37], we obtain a large deviation principle for quadratic forms of Gaussian stationary processes. It is established by the conjunction of a result of Roch and Silbermann on the spectrum of products of Toeplitz matrices together with the analysis of large deviations carried out by Gamboa, Rouault and the first author. An alternative proof of the needed result on Toeplitz matrices, based on semi-classical analysis, is also provided.

5.8. Sharp large deviations for the non-stationary Ornstein-Uhlenbeck process

For the Ornstein-Uhlenbeck process, the asymptotic behavior of the maximum likelihood estimator of the drift parameter is totally different in the stable, unstable, and explosive cases. Notwithstanding of this trichotomy, we investigate sharp large deviation principles for this estimator in the three situations. In the explosive case, we exhibit in [38] a very unusual rate function with a shaped flat valley and an abrupt discontinuity point at its minimum.

5.9. Markovian quadratic and superquadratic BSDEs with an unbounded terminal condition

The work in [43] deals with the existence and the uniqueness of solutions to quadratic and superquadratic Markovian backward stochastic differential equations (BSDEs for short) with an unbounded terminal condition. Our results are deeply linked with a strong a priori estimate on Z that takes advantage of the Markovian framework. This estimate allows us to prove the existence of a viscosity solution to a semilinear parabolic partial differential equation with nonlinearity having quadratic or superquadratic growth in the gradient of the solution. This estimate also allows us to give explicit convergence rates for time approximation of quadratic or superquadratic Markovian BSDEs.

5.10. New approach on recursive and non-recursive SIR

In [3], we consider a semiparametric single index regression model involving a p -dimensional quantitative covariable x and a real dependent variable y . A dimension reduction is included in this model via an index $x^T \beta$. Sliced inverse regression (SIR) is a well-known method to estimate the direction of the Euclidean parameter β which is based on a "slicing step" of y in the population and sample versions. The goal of this paper is twofold. On the one hand, we focus on a recursive version of SIR which is also suitable for multiple indices model. On the other hand, we propose a new method called SIRoneslice when the regression model is a single index model. The SIRoneslice estimator of the direction of β is based on the use of only one "optimal" slice chosen among the H slices. Then, we provide its recursive version. We give an asymptotic result for the SIRoneslice approach. Simulation study shows good numerical performances of the SIRoneslice method and clearly exhibits the main advantage of using recursive versions of the SIR and SIRoneslice methods from a computational time point of view. A real dataset is also used to illustrate the approach. The proposed methods and criterion have been implemented in R and the corresponding codes are made available.

5.11. Classification of EEG data by evolutionary algorithm for the study of vigilance states

The objective of this work [18] is to predict the state of vigilance of an individual from the study of its brain activity (EEG signals). The variable to predict is binary (alertness "normal" or "relaxed"). EEG of 44 participants in both states (88 records) were collected with a helmet with 58 electrodes. After a pretreatment step and data validation, a test called "test slope" was chosen. The usual methods of supervised classification (k nearest neighbors, binary classification trees, random forests, and discriminant sparse PLS) were used to provide predictions of the state of participants. The test was then refined using a genetic algorithm, which has built a reliable model (average true classification rate by using CART equal to 86.68 +/- 1.87%) and to select an electrode from the initial 58.

5.12. Genetic Programming

Recently, it has been stated that the complexity of a solution is a good indicator of the amount of overfitting it incurs. However, measuring the complexity of a program, in Genetic Programming, is not a trivial task. In [22], we study the functional complexity and how it relates with overfitting on symbolic regression problems. We consider two measures of complexity, Slope-based Functional Complexity, inspired by the concept of curvature, and Regularity-based Functional Complexity based on the concept of Holderian regularity. In general, both complexity measures appear to be poor indicators of program overfitting. However, results suggest that Regularity-based Functional Complexity could provide a good indication of overfitting in extreme cases.

During the development of applied systems, an important problem that must be addressed is that of choosing the correct tools for a given domain or scenario. This general task has been addressed by the genetic programming (GP) community by attempting to determine the intrinsic difficulty that a problem poses for a GP search. In [21], we present an approach to predict the performance of GP applied to data classification, one of the most common problems in computer science. The novelty of the proposal is to extract statistical descriptors and complexity descriptors of the problem data, and from these estimate the expected performance of a GP classifier. We derive two types of predictive models: linear regression models and symbolic regression models evolved with GP. The experimental results show that both approaches provide good estimates of classifier performance, using synthetic and real-world problems for validation. In conclusion, this paper shows that it is possible to accurately predict the expected performance of a GP classifier using a set of descriptors that characterize the problem data.

The analysis of image regularity using Holder exponents can be used to characterize singular structures contained within an image, and provide a compact description of local shape and appearance. However, estimating the Holder exponent is not a trivial task and current methods tend to be slow and complex. Therefore, the goal in [35] is to automatically synthesize image operators that can be used to estimate the Holder regularity of an image. We pose this task as an optimization problem and use Genetic Programming

(GP) to search for operators that can approximate a traditional estimator, the oscillations method. In our experiments, GP was able to evolve estimators that achieve a low error and a high correlation with the ground truth estimation. Furthermore, most of the GP estimators are faster than the traditional approaches, in some cases their runtime is orders of magnitude smaller. This result allowed us to implement a real-time estimation of the Holder exponent on a live video signal, the first such implementation in current literature. Moreover, the evolved estimators are used to generate local descriptors of salient image regions, a task for which we obtain a stable and robust matching that is comparable with state-of-the-art methods. In conclusion, the evolved estimators produced by GP could help expand the application domain of Holderian regularity within the fields of image analysis and signal processing.

5.13. Relevance of the Holderian regularity-based interpolation for range-Doppler ISAR image post-processing

In ISAR processing, post-processing of the range Doppler image is useful to help the practitioner for ship recognition. Among the image post-processing tools, interpolation methods can be of interest especially when zooming. In [19], we study the relevance of the Holderian regularity-based interpolation. In that case, interpolating consists in adding a new scale in the wavelet transform and the new wavelet coefficients can be estimated from others. In the original method, initially proposed by two of the authors, the image is first interpolated along the rows and then along the columns. Concerning the diagonal pixels, they are estimated as the mean of the adjacent original and interpolated pixels. Here, we propose a variant where the diagonal pixels are estimated by taking into account the local orientation of the image. It has the advantage of conserving local regularity on all interpolated pixels of the image. A comparative study on synthetic data and real range-Doppler images is then carried out with alternative interpolation techniques such as the linear interpolation, the bicubic one, the nearest neighbour interpolation, etc. The simulation results confirm the effectiveness of the approach.

5.14. On-line changepoint detection and parameter estimation with application to genomic data

We propose in [10] an efficient on-line changepoint detection algorithm for an important class of Bayesian product partition models. The algorithm allows to estimate jointly on-line the static parameters of the model using a recursive maximum likelihood estimation strategy. This particle filter type algorithm has a computational complexity which scales linearly both in the number of data and the number of particles. We demonstrate our methodology on a synthetic and two real world datasets from RNA transcript analysis. On simulated data, it is shown that our approach outperforms standard techniques used in this context and hence has the potential to detect novel RNA transcripts.

5.15. Bayesian Sparsity-Path-Analysis of Genetic Association Signal using Generalized t Priors

In [17], we explore the use of generalized t priors on regression coefficients to help understand the nature of association signal within "hit regions" of genome-wide association studies. The particular generalized t distribution we adopt is a Student distribution on the absolute value of its argument. For low degrees of freedom we show that the generalized t exhibits 'sparsity-prior' properties with some attractive features over other common forms of sparse priors and includes the well known double-exponential distribution as the degrees of freedom tends to infinity. We pay particular attention to graphical representations of posterior statistics obtained from sparsity-path-analysis (SPA) where we sweep over the setting of the scale (shrinkage / precision) parameter in the prior to explore the space of posterior models obtained over a range of complexities, from very sparse models with all coefficient distributions heavily concentrated around zero, to models with diffuse priors and coefficients distributed around their maximum likelihood estimates. The SPA plots are akin to LASSO plots of maximum a posteriori (MAP) estimates but they characterize the complete marginal posterior distributions of the coefficients plotted as a function of the precision of the prior. Generating

posterior distributions over a range of prior precisions is computationally challenging but naturally amenable to sequential Monte Carlo (SMC) algorithms indexed on the scale parameter. We show how SMC simulation on graphic-processing-units (GPUs) provides very efficient inference for SPA. We also present a scale-mixture representation of the generalized t prior that leads to an EM algorithm to obtain MAP estimates should only these be required.

6. Contracts and Grants with Industry

6.1. Contract with EDF

The objective of this contract (2009-2011) between the teams ALEA and CQFD and EDF, is to develop algorithms for the recursive prediction of the electricity consumption. The team will organize a workshop on this subject in Institut Henri Poincaré.

6.2. Contract with EDF

The objective of this contract is to develop particle algorithms for the pricing of American-style options [33].

6.3. Contract with CEA

The objective of this contract (2010-2011) is to propose algorithms for the estimation of uncertainties in laser experiments [30], [31].

7. Partnerships and Cooperations

7.1. Regional Initiatives

The project PSI (Psychology and sounds interactions), headed by P. Legrand received a grant by the region Aquitaine for a PhD thesis on "Dimension reduction in supervised learning. Application to the study of brain activity".

7.2. National Initiatives

7.2.1. ANR Propagation (2010-2012)

To combat dramatic event such as happened in Bombay last year (coming from the sea, a terrorist commando killed more than 200 peoples in Bombay city), authorities are decided to deploy efficient sea surveillance system to protect coastal zone including sensitive infrastructures often in vicinity of important cities.

Regulation on frequencies allocation and on coastal constructions is strong constraint to be taken into account to install technical capabilities to permanently survey vulnerable littoral zones. For example, new active sensor shall be frequencies compatible within numerous existing ones in inhabited region. In this context to perform coastal surveillance, attractive solution is to deploy passive sensors networks because:

- Not necessarily compatible within existing active sensors network.
- Provide large possibilities to install the passive sensors, because, it is not needed to be on the shoreline, but can be deployed inside the territory. Such as facility offers more potential sites and then, to optimise the deployment for optimal coverage of the sensitive zone.
- Is totally undetectable by external technical means in hand of structured criminal organisations.

For these objectives, the PROPAGATION project will study, develop and experiment a demonstrator to carry out maritime traffic picture from a set of passive sensors: passive radar, AIS and optronic cameras deployed over a coastal site. This is a joint ANR project with DCNS, Thalès, Ecomer and Exavisio, **accepted in 2009**.

7.2.2. Project PEPH

This is an interdisciplinary exploratory research project, between Institut de Mathématiques de Bordeaux and Laboratory Ecologie & Evolution, UMR 7625 CNRS-UMPC-ENS (responsible: **B. Cazelles**). The objective of this project on the dynamics of epidemic diseases characterized by multiple strains of pathogens, is to use the competencies of the ALEA team to get efficient Bayesian optimization techniques. An opening **workshop** on stochastic models and bayesian inference in epidemiology has been organized in Bordeaux in November 2011.

7.3. European Initiatives

7.3.1. Major European Organizations with which you have followed Collaborations

Partner 1: Oxford University, Department of Statistics (UK)

Interacting Particle Systems

Bayesian nonparametrics

Partner 2: Imperial College, Department of Statistics (UK)

Interacting Particle Systems

7.4. International Initiatives

7.4.1. Visits of International Scientists

The following researchers visited the Team ALEA during 2011: M. Ludkovski (Univ. UCSB), A. Doucet (Univ. Oxford), C. Holmes (Oxford), C. Archambeau (Xerox), N. Whiteley (Univ. Bristol), S. Singh (Cambridge), L. Bornn (UBC), Leonardo Trujillo (Cicese).

8. Dissemination

8.1. Animation of the scientific community

P. Del Moral is currently associate editor/editor for the following journals

- Chief editor : **ESAIM: Proceedings** since 2006.
- Associate editor : **Applied Mathematics and Optimization** since 2009.
- Associate editor **Revista de Matemática: Teoría y Aplicaciones**, since 2009.
- Associate editor : **Stochastic Analysis and Applications** since 2001.

P. Del Moral participated to the following committees

- Scientific expert for the PES research grant selection committee of the French Ministry of Research in 2010, section CNU 25 and 26.
- Responsible with Xavier Warin (EDF R&D Clamart) of the theme *modélisation stochastique et incertitude*, of the strategic action EDF-INRIA since 2010.

P. Del Moral co-organized an interdisciplinary **workshop on stochastic models and Bayesian inference in epidemiology** in Bordeaux.

B. Bercu is responsible of the thematic group **MAS** (Modélisation Aléatoire et Statistique) at SMAI.

B. Bercu is an assistant director of the Institut de Mathématiques de Bordeaux (IMB). He is also a member of the IMB council and the UFR council of the University of Bordeaux. He is a member of the CNU section 26.

- B. Bercu is co-responsible of the specialty "Modélisation Statistique et Stochastique" of the Master MIMSE.
- F. Caron, P. Legrand and P. Del Moral co-organized the [Machine Learning Summer School 2011](#), organized near Bordeaux in September 4-17, 2011.
- F. Caron gave a practical session on parametric and nonparametric Bayesian clustering at the [Machine Learning Summer School 2011](#) in Bordeaux.
- F. Caron was in the senior program committee of the Fourteenth International Conference on Artificial Intelligence and Statistics ([AISTATS 2011](#)).
- F. Caron was in the program committee of the [NIPS workshop on Choice Models and Preference Learning](#).
- F. Caron was a reviewer for the following international machine learning conferences and journals this year: International Conference on Machine Learning ([ICML 2011](#)), Neural Information Processing Systems ([NIPS 2011](#)), Journal of the Royal Statistical Society B, Bayesian Analysis, IEEE Transactions on Signal Processing, Statistics and Computing, Journal of Machine Learning Research, Automatica.
- P. Legrand was a reviewer for the following international conferences and journals this year: Signal Processing, [Evolve 2011](#), [EA 2011](#).
- P. Legrand and P. Del Moral were in the organizing committee of the international conference [Evolve 2011](#).
- P. Legrand was in the organizing committee of the French summer school [EA 2011](#) and the international conference [EA 2011](#).

8.2. Teaching

- B. Bercu is teaching the following courses (142 hours)
- Licence: Mathématiques générales, Analyse et Algèbre SVE, 36h, L1, University of Bordeaux, France
 - Master: Séries Chronologiques, 48h, M2, University of Bordeaux, France
 - Master: Processus aléatoires à temps discret, Martingales, 30h, M1, University of Bordeaux, France
 - Master: Probabilités, 30h, L3, University of Bordeaux, France
- F. Caron is teaching the following courses (50 hours)
- Master : [Unsupervised Learning](#), 25 hours, M2, University of Bordeaux, France
 - Master : Bayesian Methods, 13 hours, M2, University of Bordeaux, France
 - Master: Projet Informatique, 12 hours, M2, University of Bordeaux, France
- P. Del Moral gives the following courses
- **Since september 2011** : [Professeur chargé de cours Polytechnique, CMAP](#) (58h).
Travaux dirigés/Petites classes :
 - 1) les méthodes stochastiques et les méthodes de Monte Carlo
 - 2) Les modèles aléatoires en écologie et évolution.
- P. Legrand is teaching the following courses (244 hours)
- Licence: Analyse, 32h, L1, University of Bordeaux, France
 - Licence: Mathématiques générales, 72h, L1, University of Bordeaux, France
 - Licence: Informatique pour les mathématiques, 72h, L1, University of Bordeaux, France
 - Licence: Complément d'algèbre, 72h, L2, University of Bordeaux, France

A.Richou is teaching the following courses (128 hours)

Master: Probabilité, 32h, M1, University of Bordeaux 1, France

Licence: Probabilités et Statistiques, 32h, L3, University of Bordeaux 1, France

Licence: Probabilité et Statistiques, L3, 32h, University of Bordeaux 1, France

Licence: Probabilité et Statistiques, L1, 32h, University of Bordeaux 1, France

PhD & HdR:

PhD : Michele Pace, Stochastic models and methods for multi-object tracking [1], University of Bordeaux, July 13, 2011, supervised by P. Del Moral and F. Caron

9. Bibliography

Publications of the year

Doctoral Dissertations and Habilitation Theses

- [1] M. PACE. *Stochastic models and methods for multi-object tracking*, Université Sciences et Technologies - Bordeaux I, July 2011, <http://hal.inria.fr/tel-00651396/en>.

Articles in International Peer-Reviewed Journal

- [2] B. BERCU, L. COUTIN, N. SAVY. *Sharp large deviations for the fractional Ornstein-Uhlenbeck process*, in "SIAM Theory of Probability and its Applications", 2011, vol. 55, n^o 4, p. 575-610, <http://hal.inria.fr/hal-00386239/en>.
- [3] B. BERCU, T. MONG NGOC NGUYEN, J. SARACCO. *A new approach on recursive and non-recursive SIR methods*, in "Journal of the Korean Statistical Society", 2012, to appear [DOI : 10.1016/J.JKSS.2011.05.005], <http://hal.inria.fr/hal-00642653/en>.
- [4] L. BORNN, F. CARON. *Bayesian clustering in decomposable graphs*, in "Bayesian Analysis", 2011, vol. 6, n^o 4, <http://hal.inria.fr/inria-00487168/en>.
- [5] F. CARON, F. D'ERRICO, P. DEL MORAL, F. SANTOS, J. ZILHÃO. *The Reality of Neandertal Symbolic Behavior at the Grotte du Renne, Arcy-sur-Cure, France*, in "PLoS ONE", 2011, vol. 6, n^o 6 [DOI : 10.1371/JOURNAL.PONE.0021545], <http://hal.inria.fr/inria-00598460/en>.
- [6] F. CARON, P. DEL MORAL, A. DOUCET, M. PACE. *On the Conditional Distributions of Spatial Point Processes*, in "Advances in Applied Probability", 2011, vol. 43, n^o 2, 16, <http://hal.inria.fr/inria-00464127/en>.
- [7] F. CARON, P. DEL MORAL, A. DOUCET, M. PACE. *Particle approximations of a class of branching distribution flows arising in multi-target tracking*, in "SIAM Journal on Control and Optimization", 2011, vol. 49, n^o 4, p. 1766-1792 [DOI : 10.1137/100788987], <http://hal.inria.fr/inria-00464130/en>.
- [8] F. CARON, P. DEL MORAL, M. PACE, B.-N. VO. *On the Stability and the Approximation of Branching Distribution Flows, with Applications to Nonlinear Multiple Target Filtering*, in "Stochastic Analysis and Applications", 2011, vol. 29, n^o 6 [DOI : 10.1080/07362994.2011.598797], <http://hal.inria.fr/inria-00516507/en>.

- [9] F. CARON, A. DOUCET. *Efficient Bayesian Inference for Generalized Bradley-Terry Models*, in "Journal of Computational and Graphical Statistics", 2011, 21 p., to appear, <http://hal.inria.fr/inria-00533638/en>.
- [10] F. CARON, A. DOUCET, R. GOTTARDO. *On-line changepoint detection and parameter estimation with application to genomic data*, in "Statistics and Computing", 2012, vol. 22, n^o 2, p. 579-595 [DOI : 10.1007/s11222-011-9248-x], <http://hal.inria.fr/inria-00577217/en>.
- [11] F. CÉROU, P. DEL MORAL, T. FURON, A. GUYADER. *Sequential Monte Carlo for rare event estimation*, in "Statistics and Computing", April 2011, p. 1-14 [DOI : 10.1007/s11222-011-9231-6], <http://hal.inria.fr/inria-00584352/en>.
- [12] P. DEL MORAL, A. DOUCET, A. JASRA, C. ANDRIEU. *On nonlinear Markov chain Monte Carlo via Self-interacting approximations.*, in "Bernoulli", 2011, vol. 47, n^o 3, p. 629-649, <http://hal.inria.fr/inria-00533437/en>.
- [13] P. DEL MORAL, A. DOUCET, A. JASRA. *An Adaptive sequential Monte Carlo method for approximate bayesian computation*, in "Statistics and Computing", 2011, <http://hal.inria.fr/hal-00641462/en>.
- [14] P. DEL MORAL, A. DOUCET, A. JASRA. *On adaptive resampling procedures for sequential Monte Carlo methods*, in "Bernoulli", 2011, <http://hal.inria.fr/hal-00641134/en>.
- [15] P. DEL MORAL, P. HU, N. OUDJANE, B. RÉMILLARD. *On the Robustness of the Snell envelope*, in "SIAM Journal on Financial Mathematics", 2011, vol. 2, p. 951-997, <http://hal.inria.fr/hal-00641452/en>.
- [16] P. DEL MORAL, S. RUBENTHALER, N. CHOPIN. *Stability of Feynman-Kac formulae with path-dependent potentials*, in "Stochastic Processes and their Applications", 2011, vol. 121, p. 38-60, <http://hal.inria.fr/inria-00533438/en>.
- [17] A. LEE, F. CARON, A. DOUCET, C. HOLMES. *Bayesian Sparsity-Path-Analysis of Genetic Association Signal using Generalized t Priors*, in "Statistical Applications in Genetics and Molecular Biology", 2011, to appear, <http://hal.inria.fr/inria-00598490/en>.

Articles in National Peer-Reviewed Journal

- [18] L. VEZARD, P. LEGRAND, M. CHAVENT, F. FAITA-AINSEBA, J. CLAUZEL. *Classification de données EEG par algorithme évolutionnaire pour l'étude d'états de vigilance*, in "RNTI", 2012, <http://hal.inria.fr/hal-00643438/en>.

International Conferences with Proceedings

- [19] V. CORRETTIA, P. LEGRAND, E. GRIVEL, J. LÉVY-VEHEL. *Relevance of the Hölderian regularity-based interpolation for range-Doppler ISAR image post-processing.*, in "RADAR 2011: 6th International Conference on Radar", Chengdu, China, 2011, <http://hal.inria.fr/hal-00643369/en>.
- [20] M. PACE, H. ZHANG. *Grid based PHD filtering by Fast Fourier Transform*, in "14th International Conference on Fusion", Chicago, United States, July 2011, <http://hal.inria.fr/hal-00648109/en>.
- [21] L. TRUJILLO, Y. MARTINEZ, E. GALVAN-LOPEZ, P. LEGRAND. *Predicting Problem Difficulty for Genetic Programming Applied to Data Classification*, in "Gecco 2011", Dublin, Ireland, N. KRASNOGOR (editor),

ACM New York, NY, USA ©2011, 2011, p. 1355-1362 [DOI : 10.1145/2001576.2001759], <http://hal.inria.fr/hal-00643358/en>.

- [22] L. TRUJILLO, S. SILVA, P. LEGRAND, L. VANNESCHI. *An empirical study of functional complexity as an indicator of overfitting in Genetic Programming*, in "EuroGP", Turin, Italy, 2011, p. 262-273 [DOI : 10.1007/978-3-642-20407-4_23], <http://hal.inria.fr/hal-00642530/en>.

Conferences without Proceedings

- [23] L. VEZARD, M. CHAVENT, P. LEGRAND, F. FAITA-AINSEBA, J. CLAUZEL. *Caractérisation d'états psychophysiologiques par classification de signaux EEG. Intégration de ces résultats dans le projet PSI.*, in "Journée AFIM : électroencéphalographie et composition musicale", Talence, France, June 2011, <http://hal.inria.fr/hal-00649520/en>.

Scientific Books (or Scientific Book chapters)

- [24] R. CARMONA, P. DEL MORAL, P. HU, N. OUDJANE. *Numerical methods in finance*, Proceedings in mathematics, Springer, 2011, <http://hal.inria.fr/hal-00641465/en>.
- [25] P. DEL MORAL, A. DOUCET. *Sequential Monte Carlo & genetic particle models. Theory and practice.*, Chapman & Hall, 2012, <http://hal.inria.fr/hal-00641130/en>.
- [26] P. DEL MORAL, P. HU, N. OUDJANE, R. CARMONA. *An Introduction to particle methods in finance*, in "Numerical methods in finance", R. CARMONA, P. D. MORAL, P. HU, N. OUDJANE (editors), Proceedings in mathematics, Springer, 2011, <http://hal.inria.fr/hal-00641132/en>.
- [27] P. DEL MORAL, B. RÉMILLARD, S. RUBENTHALER. *Monte Carlo approximations of american option that preserve monotonicity and convexity*, in "Numerical methods in finance", R. CARMONA, P. D. MORAL, P. HU, N. OUDJANE (editors), Springer, 2011, vol. Proceedings in mathematics, <http://hal.inria.fr/hal-00641133/en>.
- [28] E. TANTAR, A. TANTAR, P. DEL MORAL, C. COELLO, P. LEGRAND, O. SCHUTZE, P. BOUVRY. *Evolve : A bridge between probability, set oriented numerics and evolutionary computation*, Springer, 2011, to appear, <http://hal.inria.fr/hal-00641466/en>.

Books or Proceedings Editing

- [29] J.-K. HAO, P. LEGRAND, P. COLLET, N. MONMARCHÉ, E. LUTTON, M. SCHOENAUER (editors). *Artificial Evolution 2011*, Springer, 2012, to appear, <http://hal.inria.fr/hal-00643404/en>.

Research Reports

- [30] B. BERCU, F. CARON, P. DEL MORAL. *Prise en compte des incertitudes dans la déconvolution de signaux. Lot 2: estimation du mesurande.*, INRIA, May 2011, n° 2011-CEA-01, <http://hal.inria.fr/hal-00650875/en>.
- [31] B. BERCU, F. CARON, P. DEL MORAL. *Prise en compte des incertitudes dans la déconvolution de signaux. Lot 3: estimation de l'émissivité spectrale du rayonnement incident.*, INRIA, October 2011, n° 2011-CEA-02, <http://hal.inria.fr/hal-00650878/en>.

- [32] P. DEL MORAL, A. DOUCET, S. S. SINGH. *Uniform stability of a particle approximation of the optimal filter derivative*, Cambridge University, 2011, Technical Report CUED/F-INFENG/TR 668- Cambridge University Engineering Department, <http://hal.inria.fr/hal-00642618/en>.
- [33] P. DEL MORAL, P. HU, N. OUDJANE. *Snell envelope with small probability criteria*, INRIA, November 2011, n^o RR-7360, Submitted, <http://hal.inria.fr/inria-00507794/en>.
- [34] P. DEL MORAL, P. HU, L. WU. *On the concentration properties of Interacting particle processes*, INRIA, July 2011, n^o RR-7677, <http://hal.inria.fr/inria-00607684/en>.
- [35] L. TRUJILLO, P. LEGRAND, G. OLAGUE, J. LÉVY-VEHEL. *Evolving Estimators of the Pointwise Holder Exponent with Genetic Programming*, INRIA, 2011, Submitted, <http://hal.inria.fr/hal-00643387/en>.
- [36] M. ZGHAL, L. MÉVEL, P. DEL MORAL. *Interacting Kalman filter for modal identification of time varying systems*, INRIA Rennes, 2011, article soumis à MSSP Special Issue on the Identification of Time Varying Structures and Systems, <http://hal.inria.fr/hal-00641470/en>.

Other Publications

- [37] B. BERCU, J.-F. BONY, V. BRUNEAU. *Large deviations for Gaussian stationary processes and semi-classical analysis*, 2011, <http://hal.inria.fr/hal-00642658/en>.
- [38] B. BERCU, L. COUTIN, N. SAVY. *Sharp large deviations for the non-stationary Ornstein-Uhlenbeck process*, 2011, <http://hal.inria.fr/hal-00645074/en>.
- [39] B. BERCU, P. FRAYSSE. *A Robbins-Monro procedure for estimation in semiparametric regression models*, 2011, <http://hal.inria.fr/hal-00551832/en>.
- [40] B. BERCU, F. PROIA. *A sharp analysis on the asymptotic behavior of the Durbin-Watson statistic for the first-order autoregressive process*, 2011, <http://hal.inria.fr/hal-00642634/en>.
- [41] L. BORNN, P. JACOB, P. DEL MORAL, A. DOUCET. *An Adaptive Interacting Wang-Landau Algorithm for Automatic Density Exploration*, 2011, <http://hal.inria.fr/hal-00634211/en>.
- [42] A. BUDHIRAJA, P. DEL MORAL, S. RUBENTHALER. *Discrete Time Markovian Agents Interacting Through a Potential*, 2011, 21 pages, <http://hal.inria.fr/hal-00601151/en>.
- [43] A. RICHO. *Markovian quadratic and superquadratic BSDEs with an unbounded terminal condition*, 2011, <http://hal.inria.fr/hal-00643198/en>.

References in notes

- [44] N. BARTOLI, P. DEL MORAL. *Simulation and Stochastic Algorithms (An introduction with applications)*, Cépaduès Editions, 2001.
- [45] P. DEL MORAL. *Feynman-Kac formulae. Genealogical and interacting particle systems, with applications*, Springer Verlag New York, Series: Probability and its Applications, 2004.

- [46] P. DEL MORAL, L. MICLO. *Genealogies and Increasing Propagations of Chaos for Feynman-Kac and Genetic Models*, in "Annals of Applied Probability", 2001, vol. 11, n^o 4, p. 1166-1198.
- [47] P. DEL MORAL, L. MICLO. *On convergence of chains with time empirical self-interactions*, in "Proc. Royal Soc. Lond. A.", 2003, vol. 460, p. 325-346.
- [48] P. DEL MORAL, L. MICLO. *Self Interacting Markov chains*, in "Stochastic Analysis and Applications", 2006, vol. 24, n^o 3, p. 615-660.
- [49] A. JASRA, C. ANDRIEU, A. DOUCET, P. DEL MORAL. *Non-Linear Markov chain Monte Carlo via Self Interacting Approximations*, 2007, Univ. Bristol Preprints.
- [50] A. JASRA, C. ANDRIEU, A. DOUCET, P. DEL MORAL. *A Note on Convergence of the Equi-energy Sampler*, in "Journal Stochastic Analysis and its Applications", 2008.