IN PARTNERSHIP WITH:
CNRS
Université de Bordeaux

Activity Report 2017

Project-Team CQFD

Quality control and dynamic reliability

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

RESEARCH CENTER
Bordeaux - Sud-Ouest

THEME
Stochastic approaches

Table of contents

Project-Team CQFD

Creation of the Project-Team: 2009 January 01

Keywords:

Computer Science and Digital Science:

A1.1.6. - Cloud
A1.2.4. - QoS, performance evaluation
A1.3. - Distributed Systems
A3.3. - Data and knowledge analysis
A3.4.1. - Supervised learning
A3.4.2. - Unsupervised learning
A3.4.5. - Bayesian methods
A3.4.6. - Neural networks
A3.4.7. - Kernel methods
A5.9.2. - Estimation, modeling
A5.9.6. - Optimization tools
A6.1.2. - Stochastic Modeling (SPDE, SDE)
A6.1.3. - Discrete Modeling (multi-agent, people centered)
A6.2.2. - Numerical probability
A6.2.3. - Probabilistic methods
A6.2.4. - Statistical methods
A6.2.6. - Optimization
A6.4.2. - Stochastic control
A8.11. - Game Theory
A9.2. - Machine learning
A9.6. - Decision support

Other Research Topics and Application Domains:

B2.2.4. - Infectious diseases, Virology
B2.6.1. - Brain imaging
B5.9. - Industrial maintenance
B6.2. - Network technologies
B6.3.3. - Network Management
B6.5. - Information systems
B9.2.3. - Video games
B9.4.2. - Mathematics

1. Personnel

Research Scientists

Jonatha Anselmi [Inria, Researcher]
Pierre Del Moral [Inria, Senior Researcher, HDR]

Faculty Members

Francois Dufour [Team leader, Institut National Polytechnique de Bordeaux, Professor, HDR]
Marie Chavent [Univ de Bordeaux, Associate Professor, HDR]
Alexandre Genadot [Univ de Bordeaux, Associate Professor]
Pierrick Legrand [Univ de Bordeaux, Associate Professor]
Jerome Saracco [Institut National Polytechnique de Bordeaux, Professor, HDR]
Huilong Zhang [Univ de Bordeaux, Associate Professor]

PhD Students

Alizée Geeraert [Thales, until Aug 2017]
Jessica Sodjo [Inria, until Aug 2017]
Tiffany Cherchi [Montpellier University]
Maud Joubaud [Montpellier University]
Chloé Pasin [Bordeaux University]

Administrative Assistant

Chrystel Plumejeau [Inria]

Visiting Scientist

Oswaldo Luiz Do Valle Costa [Escola Politécnica da Universidade de São Paulo, Brazil]

External Collaborator

Benoite de Saporta [Univ Montpellier II (sciences et techniques du Languedoc), HDR]

2. Overall Objectives

2.1. Presentation

The core component of our scientific agenda focuses on the development of statistical and probabilistic methods for the modeling and the optimization of complex systems. These systems require dynamic and stochastic mathematical representations with discrete and/or continuous variables. Their complexity poses genuine scientific challenges that can be addressed through complementary approaches and methodologies:

- *Modeling:* design and analysis of realistic and tractable models for such complex real-life systems taking into account various probabilistic phenomena;
- *Estimation:* developing theoretical and computational methods in order to estimate the parameters of the model and to evaluate the performance of the system;
- *Control:* developing theoretical and numerical control tools to optimize the performance.

These three approaches are strongly connected and the most important feature of the team is to consider these topics as a whole. This enables the team to deal with real industrial problems in several contexts such as biology, production planning, trajectory generation and tracking, performance and reliability.

3. Research Program

3.1. Introduction

The scientific objectives of the team are to provide mathematical tools for modeling and optimization of complex systems. These systems require mathematical representations which are in essence dynamic, multi-model and stochastic. This increasing complexity poses genuine scientific challenges in the domain of modeling and optimization. More precisely, our research activities are focused on stochastic optimization and (parametric, semi-parametric, multidimensional) statistics which are complementary and interlinked topics. It is essential to develop simultaneously statistical methods for the estimation and control methods for the optimization of the models.

3.2. Main research topics

Stochastic modeling: Markov chain, Piecewise Deterministic Markov Processes (PDMP), Markov Decision Processes (MDP).
The mathematical representation of complex systems is a preliminary step to our final goal corresponding to the optimization of its performance. The team CQFD focuses on two complementary types of approaches. The first approach is based on mathematical representations built upon physical models where the dynamic of the real system is described by *stochastic processes*. The second one consists in studying the modeling issue in an abstract framework where the real system is considered as black-box. In this context, the outputs of the system are related to its inputs through a *statistical model*. Regarding stochastic processes, the team studies Piecewise Deterministic Markov Processes (PDMPs) and Markov Decision Processes (MDPs). These two classes of Markov processes form general families of controlled stochastic models suitable for the design of sequential decision-making problems. They appear in many fields such as biology, engineering, computer science, economics, operations research and provide powerful classes of processes for the modeling of complex systems. Our contribution to this topic consists in expressing real-life industrial problems into these mathematical frameworks. Regarding statistical methods, the team works on dimension reduction models. They provide a way to understand and visualize the structure of complex data sets. Furthermore, they are important tools in several different areas such as data analysis and machine learning, and appear in many applications such as biology, genetics, environment and recommendation systems. Our contribution to this topic consists in studying semiparametric modeling which combines the advantages of parametric and nonparametric models.

Estimation methods: estimation for PDMP; estimation in non- and semi- parametric regression modeling.
To the best of our knowledge, there does not exist any general theory for the problems of estimating parameters of PDMPs although there already exist a large number of tools for sub-classes of PDMPs such as point processes and marked point processes. To fill the gap between these specific models and the general class of PDMPs, new theoretical and mathematical developments will be on the agenda of the whole team. In the framework of non-parametric regression or quantile regression, we focus on kernel estimators or kernel local linear estimators for complete data or censored data. New strategies for estimating semi-parametric models via recursive estimation procedures have also received an increasing interest recently. The advantage of the recursive estimation approach is to take into account the successive arrivals of the information and to refine, step after step, the implemented estimation algorithms. These recursive methods do require restarting calculation of parameter estimation from scratch when new data are added to the base. The idea is to use only the previous estimations and the new data to refresh the estimation. The gain in time could be very interesting and there are many applications of such approaches.

Dimension reduction: dimension-reduction via SIR and related methods, dimension-reduction via multidimensional and classification methods.
Most of the dimension reduction approaches seek for lower dimensional subspaces minimizing the loss of some statistical information. This can be achieved in modeling framework or in exploratory data analysis context.
In modeling framework we focus our attention on semi-parametric models in order to conjugate the advantages of parametric and nonparametric modeling. On the one hand, the parametric part of the model allows a suitable interpretation for the user. On the other hand, the functional part of the model offers a lot of flexibility. In this project, we are especially interested in the semi-parametric regression model $Y = f(X'\theta) + \varepsilon$, the unknown parameter θ belongs to $\mathbb{R}^p$ for a single index model, or is such that $\theta = [\theta_1, \cdots, \theta_d]$ (where each θ_k belongs to $\mathbb{R}^p$ and $d \leq p$ for a multiple indices model), the noise ε is a random error with unknown distribution, and the link function f is an unknown real valued function. Another way to see this model is the following: the variables X and Y are independent given $X'\theta$. In our semi-parametric framework, the main objectives are to estimate the parametric part θ as well as the nonparametric part which can be the link function f, the conditional distribution function of Y given X or the conditional quantile q_α. In order to estimate the dimension reduction parameter θ we focus on the Sliced Inverse Regression (SIR) method which has been introduced by Li [53] and Duan and Li [51].

Methods of dimension reduction are also important tools in the field of data analysis, data mining and machine learning.They provide a way to understand and visualize the structure of complex data sets.Traditional methods among others are principal component analysis for quantitative variables or multiple component analysis for qualitative variables. New techniques have also been proposed to address these challenging tasks involving many irrelevant and redundant variables and often comparably few observation units. In this context, we focus on the problem of synthetic variables construction, whose goals include increasing the predictor performance and building more compact variables subsets. Clustering of variables is used for feature construction. The idea is to replace a group of ”similar” variables by a cluster centroid, which becomes a feature. The most popular algorithms include K-means and hierarchical clustering. For a review, see, e.g., the textbook of Duda [52].

Stochastic control: optimal stopping, impulse control, continuous control, linear programming.
The main objective is to develop *approximation techniques* to provide quasi-optimal feasible solutions and to derive *optimality results* for control problems related to MDPs and PDMPs:

- *Approximation techniques*. The analysis and the resolution of such decision models mainly rely on the maximum principle and/or the dynamic/linear programming techniques together with their various extensions such as the value iteration (VIA) and the policy iteration (PIA) algorithm. However, it is well known that these approaches are hardly applicable in practice and suffer from the so-called *curse of dimensionality*. Hence, solving numerically a PDMP or an MDP is a difficult and important challenge. Our goal is to obtain results which are both consistent from a theoretical point of view and computationally tractable and accurate from an application standpoint. It is important to emphasize that these research objectives were not planned in our initial 2009 program.

  Our objective is to propose approximation techniques to efficiently compute the optimal value function and to get quasi-optimal controls for different classes of constrained and unconstrained MDPs with general state/action spaces, and possibly unbounded cost function. Our approach is based on combining the linear programming formulation of an MDP with probabilistic approximation techniques related to quantization techniques and the theory of empirical processes. An other aim is to apply our methods to specific industrial applications in collaboration with industrial partners such as Airbus Defence & Space, DCNS and Thales.

  Asymptotic approximations are also developed in the context of queueing networks, a class of models where the decision policy of the underlying MDP is in some sense fixed a priori, and our main goal is to study the transient or stationary behavior of the induced Markov process. Even though the decision policy is fixed, these models usually remain intractable to solve. Given this complexity, the team has developed analyses in some limiting regime of practical interest, i.e., queueing models in the large-network, heavy-traffic, fluid or mean-field limit. This approach is helpful to obtain a simpler mathematical description of the system under investigation, which is often given in terms of ordinary differential equations or convex optimization problems.

- *Optimality results*. Our aim is to investigate new important classes of optimal stochastic control problems including constraints and combining continuous and impulse actions for MDPs and PDMPs. In this framework, our objective is to obtain different types of optimality results. For example, we intend to provide conditions to guarantee the existence and uniqueness of the optimality equation for the problem under consideration and to ensure existence of an optimal (and ϵ-optimal) control strategy. We also plan to analyze the structural properties of the optimal strategies as well as to study the associated infinite dimensional linear programming problem. These results can be seen as a first step toward the development of numerical approximation techniques in the sense described above.

4. Application Domains

4.1. Dependability and safety

Our abilities in probability and statistics apply naturally to industry, in particular in studies of dependability and safety. An illustrative example is the collaboration that started in September 2014 with with THALES Optronique. The goal of this project is the optimization of the maintenance of an onboard system equipped with a HUMS (Health Unit Monitoring Systems). The physical system under consideration is modeled by a piecewise deterministic Markov process. In the context of impulse control, we propose a dynamic maintenance policy, adapted to the state of the system and taking into account both random failures and those related to the degradation phenomenon.

The spectrum of applications of the topics that the team can address is large and can concern many other fields. Indeed non parametric and semi-parametric regression methods can be used in biometry, econometrics or engineering for instance. Gene selection from microarray data and text categorization are two typical application domains of dimension reduction among others. We had for instance the opportunity via the scientific program PRIMEQUAL to work on air quality data and to use dimension reduction techniques as principal component analysis (PCA) or positive matrix factorization (PMF) for pollution sources identification and quantization.

5. Highlights of the Year

5.1. Highlights of the Year

Pierre Del Moral is a Simons foundation CRM Professor, Montréal Math. Research Center 2017

Google scholar classic paper in Probability and Statistics (ten most-cited articles published ten years earlier): Del Moral, P., Doucet A., Jasra A.. *Sequential Monte Carlo Samplers* Journal of the Royal Statistical Society, Series B, vol. 68, no. 3, pp. 411-436 (2006).

6. New Software and Platforms

6.1. biips

Bayesian Inference with Interacting Particle Systems

FUNCTIONAL DESCRIPTION: Biips is a software platform for automatic Bayesian inference with interacting particle systems. Biips allows users to define their statistical model in the probabilistic programming BUGS language, as well as to add custom functions or samplers within this language. Then it runs sequential Monte Carlo based algorithms (particle filters, particle independent Metropolis-Hastings, particle marginal Metropolis-Hastings) in a black-box manner so that to approximate the posterior distribution of interest as well as the marginal likelihood. The software is developed in C++ with interfaces with the softwares R, Matlab and Octave.

- Participants: Adrien Todeschini and François Caron
- Contact: Adrien Todeschini
- URL: http://biips.gforge.inria.fr

6.2. PCAmixdata

KEYWORD: Statistic analysis

FUNCTIONAL DESCRIPTION: Mixed data type arise when observations are described by a mixture of numerical and categorical variables. The R package PCAmixdata extends standard multivariate analysis methods to incorporate this type of data. The key techniques included in the package are PCAmix (PCA of a mixture of numerical and categorical variables), PCArot (rotation in PCAmix) and MFAmix (multiple factor analysis with mixed data within a dataset). The MFAmix procedure handles a mixture of numerical and categorical variables within a group - something which was not possible in the standard MFA procedure. We also included techniques to project new observations onto the principal components of the three methods in the new version of the package.

- Contact: Marie Chavent
- URL: https://cran.r-project.org/web/packages/PCAmixdata/index.html

6.3. QuantifQuantile

KEYWORD: Regression
FUNCTIONAL DESCRIPTION: QuantifQuantile is an R package that allows to perform quantization-based quantile regression. The different functions of the package allow the user to construct an optimal grid of N quantizers and to estimate conditional quantiles. This estimation requires a data driven selection of the size N of the grid that is implemented in the functions. Illustration of the selection of N is available, and graphical output of the resulting estimated curves or surfaces (depending on the dimension of the covariate) is directly provided via the plot function.

- Contact: Jérôme Saracco

7. New Results

7.1. Asymptotically optimal open-loop load balancing

In many distributed computing systems, stochastically arriving jobs need to be assigned to servers with the objective of minimizing waiting times. Many existing dispatching algorithms are basically included in the SQ(d) framework: Upon arrival of a job, $d \geq 2$ servers are contacted uniformly at random to retrieve their state and then the job is routed to a server in the best observed state. One practical issue in this type of algorithm is that server states may not be observable, depending on the underlying architecture. In [3], we investigate the assignment problem in the open-loop setting where no feedback information can flow dynamically from the queues back to the controller, i.e., the queues are unobservable. This is an intractable problem, and unless particular cases are considered, the structure of an optimal policy is not known. Under mild assumptions and in a heavy-traffic many-server limiting regime, our main result proves the optimality of a subset of deterministic and periodic policies within a wide set of (open-loop) policies that can be randomized or deterministic and can be dependent on the arrival process at the controller. The limiting value of the scaled stationary mean waiting time achieved by any policy in our subset provides a simple approximation for the optimal system performance.

Author: J. Anselmi (Inria CQFD).

7.2. The economics of the cloud: price competition and congestion

The work developed in [4] proposes a model to study the interaction of price competition and congestion in the cloud computing marketplace. Specifically, we propose a three-tier market model that captures a marketplace with users purchasing services from Software-as-Service (SaaS) providers, which in turn purchase computing resources from either Provider-as-a-Service (PaaS) providers or Infrastructure-as-a-Service (IaaS) providers. Within each level, we define and characterize competitive equilibria. Further, we use these characterizations to understand the relative profitability of SaaSs and PaaSs/IaaSs, and to understand the impact of price competition on the user experienced performance, i.e., the 'price of anarchy' of the cloud marketplace. Our results highlight that both of these depend fundamentally on the degree to which congestion results from shared or dedicated resources in the cloud.

Authors: J. Anselmi (Inria CQFD), D. Ardagna, J.C.S. Lui, A. Wierman, Y. Xu and Z. Yang.

7.3. A new characterization of the jump rate for piecewise-deterministic Markov processes with discrete transitions

Piecewise-deterministic Markov processes form a general class of non-diffusion stochastic models that involve both deterministic trajectories and random jumps at random times. In [5], we state a new characterization of the jump rate of such a process with discrete transitions. We deduce from this result a nonparametric technique for estimating this feature of interest. We state the uniform convergence in probability of the estimator. The methodology is illustrated on a numerical example.

Authors: A. Genadot (Inria CQFD) and R. Azais.

7.4. Linear minimum mean square filters for Markov jump linear systems

In [9], new linear minimum mean square estimators are introduced by considering a cluster information structure in the filter design. The set of filters constructed in this way can be ordered in a lattice according to the refines of clusters of the Markov chain, including the linear Markovian estimator at one end (with only one cluster) and the Kalman filter at the other hand (with as many clusters as Markov states). The higher is the number of clusters, the heavier are pre-compuations and smaller is the estimation error, so that the cluster cardinality allows for a trade-off between performance and computational burden. In this paper we propose the estimator, give the formulas for pre-computation of gains, present some properties, and give an illustrative numerical example.

Authors: E. Costa and B. De Saporta (Inria CQFD).

7.5. Zero-sum discounted reward criterion games for piecewise deterministic Markov processes

In [10], we deal with zero-sum games with a discounted reward criterion for piecewise deterministic Markov process (PDMPs) in general Borel spaces. The two players can act on the jump rate and transition measure of the process, with the decisions being taken just after a jump of the process. The goal of this paper is to derive conditions for the existence of minâmax strategies for the infinite horizon total expected discounted reward function, which is composed of running and boundary parts. The basic idea is, by using the special features of the PDMPs, to re-write the problem via an embedded discrete-time Markov chain associated to the PDMP and re-formulate the problem as a discrete-stage zero sum game problem.

Authors: O. Costa and F. Dufour (Inria CQFD).

7.6. Optimal strategies for impulse control of piecewise deterministic Markov processes

In [11], we deal with the general discounted impulse control problem of a piecewise deterministic Markov process. We investigate a new family of optimal strategies. The construction of such strategies is explicit and only necessitates the previous knowledge of the cost of the no-impulse strategy. In particular, it does not require the resolution of auxiliary optimal stopping problem or the computation of the value function at each point of the state space. This approach is based on the iteration of a single-jump-orintervention operator associated to the piecewise deterministic Markov process.

Authors: B. De Saporta, F. Dufour and A. Geeraert. All authors are members of CQFD at Inria.

7.7. Partially observed optimal stopping problem for discrete-time Markov processes

In [12], we have investigated of a new numerical method to approximate the optimal stopping problem for a discrete-time continuous state space Markov chain under partial observations. It is based on a two-step discretization procedure based on optimal quantization. First,we discretize the state space of the unobserved variable by quantizing an underlying reference measure. Then we jointly discretize the resulting approximate filter and the observation process. We obtain a fully computable approximation of the value function with explicit error bounds for its convergence towards the true value fonction.

Authors: B. De Saporta, F. Dufour and C. Nivot. All authors are members of CQFD at Inria.

7.8. On the stability and the uniform propagation of chaos of a class of extended ensemble Kalman–Bucy filters

The result published in [15] deals with the exponential stability and the uniform propagation of chaos properties of a class of Extended Ensemble Kalman-Bucy filters with respect to the time horizon. This class of nonlinear filters can be interpreted as the conditional expectations of nonlinear McKean Vlasov type diffusions with respect to the observation process. In contrast with more conventional Langevin nonlinear drift type processes, the mean field interaction is encapsulated in the covariance matrix of the diffusion. The main results discussed in the article are quantitative estimates of the exponential stability properties of these nonlinear diffusions. These stability properties are used to derive uniform and non asymptotic estimates of the propagation of chaos properties of Extended Ensemble Kalman filters, including exponential concentration inequalities. To our knowledge these results seem to be the first results of this type for this class of nonlinear ensemble type Kalman-Bucy filters.

Authors: P. Del Moral (Inria CQFD), A. Kurtzmann and J. Tugaut.

7.9. Exponential mixing properties for time inhomogeneous diffusion processes with killing

In [16], we consider an elliptic and time-inhomogeneous diffusion process with time-periodic coefficients evolving in a bounded domain of R d with a smooth boundary. The process is killed when it hits the boundary of the domain (hard killing) or after an exponential time (soft killing) associated with some bounded rate function. The branching particle interpretation of the non absorbed diffusion again behaves as a set of interacting particles evolving in an absorbing medium. Between absorption times, the particles evolve independently one from each other according to the diffusion evolution operator; when a particle is absorbed, another selected particle splits into two offsprings. This article is concerned with the stability properties of these non absorbed processes. Under some classical ellipticity properties on the diffusion process and some mild regularity properties of the hard obstacle boundaries, we prove an uniform exponential strong mixing property of the process conditioned to not be killed. We also provide uniform estimates w.r.t. the time horizon for the interacting particle interpretation of these non-absorbed processes, yielding what seems to be the first result of this type for this class of diffusion processes evolving in soft and hard obstacles, both in homogeneous and non-homogeneous time settings.

Authors: P. Del Moral (Inria CQFD) and D. Villemonais.

7.10. Averaging for some simple constrained Markov processes

In [17], we study a class of piecewise deterministic Markov processes with underlying fast dynamic. Using a " penalty method " , an averaging result is obtained when the underlying dynamic is infinitely accelerated. The features of the averaged process, which is still a piecewise deterministic Markov process, are fully described.

Authors: A. Genadot (Inria CQFD).

7.11. Nonasymptotic analysis of adaptive and annealed Feynman–Kac particle models

Sequential and quantum Monte Carlo methods, as well as genetic type search algorithms can be interpreted as a mean field and interacting particle approximations of Feynman-Kac models in distribution spaces. The performance of these population Monte Carlo algorithms is strongly related to the stability properties of nonlinear Feynman–Kac semigroups. In [18], we analyze these models in terms of Dobrushin ergodic coefficients of the reference Markov transitions and the oscillations of the potential functions. Sufficient conditions for uniform concentration inequalities w.r.t. time are expressed explicitly in terms of these two quantities. We provide an original perturbation analysis that applies to annealed and adaptive Feynman-Kac models, yielding what seems to be the first results of this kind for these types of models. Special attention is devoted to the particular case of Boltzmann-Gibbs measures' sampling. In this context, we design an explicit way of tuning the number of Markov chain Monte Carlo iterations with temperature schedule. We also design an alternative interacting particle method based on an adaptive strategy to define the temperature increments. The theoretical analysis of the performance of this adaptive model is much more involved as both the potential functions and the reference Markov transitions now depend on the random evolution on the particle model. The nonasymptotic analysis of these complex adaptive models is an open research problem. We initiate this study with the concentration analysis of a simplified adaptive models based on reference Markov transitions that coincide with the limiting quantities, as the number of particles tends to infinity.

Authors: F. Giraud and P. Del Moral (Inria CQFD).

7.12. A comparison of fitness-case sampling methods for genetic programming

Genetic programming (GP) is an evolutionary computation paradigm for automatic program induction. GP has produced impressive results but it still needs to overcome some practical limitations, particularly its high computational cost, overfitting and excessive code growth. Recently, many researchers have proposed fitness-case sampling methods to overcome some of these problems, with mixed results in several limited tests. In [20], we present an extensive comparative study of four fitness-case sampling methods, namely: Interleaved Sampling, Random Interleaved Sampling, Lexicase Selection and Keep-Worst Interleaved Sampling. The algorithms are compared on 11 symbolic regression problems and 11 supervised classification problems, using 10 synthetic benchmarks and 12 real-world data-sets. They are evaluated based on test performance, overfitting and average program size, comparing them with a standard GP search. Comparisons are carried out using non-parametric multigroup tests and post hoc pairwise statistical tests. The experimental results suggest that fitness-case sampling methods are particularly useful for difficult real-world symbolic regression problems, improving performance, reducing overfitting and limiting code growth. On the other hand, it seems that fitness-case sampling cannot improve upon GP performance when considering supervised binary classification.

Authors: Y. Martinez, E. Naredo, L. Trujillo, P. Legrand (Inria CQFD) and U. Lopez.

7.13. Stochastic control of observer trajectories in passive tracking with acoustic signal propagation optimization

In [23], we present a numerical method which computes the optimal trajectory of a underwater vehicle subject to some mission objectives. The method is applied to a submarine whose goal is to best detect one or several targets, or/and to minimize its own detection range perceived by the other targets. The signal considered is acoustic propagation attenuation. Our approach is based on dynamic programming of a finite horizon Markov decision process. A quantization method is applied to fully discretize the problem and allows a numerically tractable solution. Different scenarios are considered. We suppose at first that the position and the velocity of the targets are known and in the second we suppose that they are unknown and estimated by a Kalman type filter in a context of bearings-only tracking.

Authors: H. Zhang (Inria CQFD), B. De Saporta (Inria CQFD), F. Dufour (Inria CQFD), D. Laneuville and A. Nègre.

7.14. Use of local Search in Genetic Programming

There are two important limitations of standard tree-based genetic programming (GP). First, GP tends to evolve unnecessarily large programs, what is referred to as bloat. Second, GP uses inefficient search operators that focus on modifying program syntax. The first problem has been studied in many works, with many bloat control proposals. Regarding the second problem, one approach is to use alternative search operators, for instance geometric semantic operators, to improve convergence. In [36], our goal is to experimentally show that both problems can be effectively addressed by incorporating a local search optimizer as an additional search operator. Using real-world problems, we show that this rather simple strategy can improve the convergence and performance of tree-based GP, while reducing program size. Given these results, a question arises: why are local search strategies so uncommon in GP? A small survey of popular GP libraries suggests to us that local search is underused in GP systems.

Authors: Leonardo Trujillo, Emigdio Z-Flores, Perla S. Juarez Smith, Pierrick Legrand (Inria CQFD), Sara Silva, Mauro Castelli, Leonardo Vanneschi, Oliver Schutze and Luis Munoz.

7.15. Hierarchical clustering with spatial constraints

In [8], we propose a Ward-like hierarchical clustering algorithm including spatial/geographical constraints. Two dissimilarity matrices D_0 and D_1 are inputted, along with a mixing parameter $\alpha \in [0,1]$. The dissimilarities can be non-Euclidean and the weights of the observations can be non-uniform. The first matrix gives the dissimilarities in the "feature space" and the second matrix gives the dissimilarities in the "constraint space". The criterion minimized at each stage is a convex combination of the homogeneity criterion calculated with D_0 and the homogeneity criterion calculated with D_1. The idea is then to determine a value of α which increases the spatial contiguity without deteriorating too much the quality of the solution based on the variables of interest i.e. those of the feature space. This procedure is illustrated on a real dataset using the R package ClustGeo.

Authors: Marie Chavent (Inria CQFD), Vanessa Kuentz, Amaury Labenne, Jérôme Saracco (Inria CQFD).

7.16. Variable importance assessment in sliced inverse regression for variable selection

In [19], we are interested in treating the relationship between a dependent variable y and a multivariate covariate x in a semiparametric regression model. Since the purpose of most social, biological, or environmental science research is the explanation, the determination of the importance of the variables is a major concern. It is a way to determine which variables are the most important when predicting y. Sliced inverse regression methods allows to reduce the space of the covariate x by estimating the directions β that form an effective dimension reduction (EDR) space. The aim of this article is to propose a computational method based on importance variable measure (only relying on the EDR space) in order to select the most useful variables. The numerical behavior of this new method, implemented in R, is studied on a simulation study. An illustration on a real data is also provided.

Authors: Ines Jlassi, Jérôme Saracco (Inria CQFD).

7.17. Group-sparse block PCA and explained variance

In [46], we address the simultaneous determination of group-sparse loadings by block optimization, and the correlated problem of defining explained variance for a set of non orthogonal components. We give in both cases a comprehensive mathematical presentation of the problem, which leads to propose i) a new formulation/algorithm for group-sparse block PCA and ii) a framework for the definition of explained variance with the analysis of five definitions. The numerical results i) confirm the superiority of block optimization over deflation for the determination of group-sparse loadings, and the importance of group information when available, and ii) show that ranking of algorithms according to explained variance is essentially independent of the definition of explained variance. These results lead to propose a new optimal variance as the definition of choice for explained variance.

Authors: Guy Chavent, Marie Chavent (Inria CQFD).

7.18. Multivariate Analysis of Mixed Data

In [47], we focus on mixed data that arise when observations are described by a mixture of numerical and categorical variables. The R package PCAmixdata extends standard multivariate analysis methods to incorporate this type of data. The key techniques/methods included in the package are principal component analysis for mixed data (PCAmix), varimax-like orthogonal rotation for PCAmix, and multiple factor analysis for mixed multi-table data. This paper gives a synthetic presentation of the three algorithms with details to help the user understand graphical and numerical outputs of the corresponding R functions. The three main methods are illustrated on a real dataset composed of four data tables characterizing living conditions in different municipalities in the Gironde region of southwest France.

Authors: Marie Chavent (Inria CQFD), Vanessa Kuentz, Amaury Labenne, Jérôme Saracco (Inria CQFD).

7.19. A Smooth Nonparametric Estimator of a Conditional Quantile

In [50], we propose a new smooth nonparametric estimator of conditional quantile of Y for a given value of X using a kernel type of estimators. A numerical study to examine the performance of our estimator as well as a theoretical asymptotic study have been conducted.

Authors: Ines Jlassi, Jérôme Saracco (Inria CQFD).

7.20. Perturbations and projections of Kalman-Bucy semigroups

The purpose of the work published in [40] is to analyse the effect of various perturbations and projections of Kalman-Bucy semigroups and Riccati equations. The original motivation was to understand the behaviour of various regulation methods used in ensemble Kalman filtering (EnKF). For example, covariance inflation-type methods (perturbations) and covariance localisation methods (projections) are commonly used in the EnKF literature to ensure well-posedness of the sample covariance (e.g. sufficient rank) and to 'move' the sample covariance closer (in some sense) to the Riccati flow of the true Kalman filter. In the limit, as the number of samples tends to infinity, these methods drive the sample covariance toward a solution of a perturbed, or projected, version of the standard (Kalman-Bucy) differential Riccati equation. The behaviour of this modified Riccati equation is investigated here. Results concerning continuity (in terms of the perturbations), boundedness, and convergence of the Riccati flow to a limit are given. In terms of the limiting filters, results characterising the error between the perturbed/projected and nominal conditional distributions are given. New projection-type models and ideas are also discussed within the EnKF framework; e.g. projections onto so-called Bose-Mesner algebras. This work is generally important in understanding the limiting bias in both the EnKF empirical mean and covariance when applying regularisation. Finally, we note the perturbation and projection models considered herein are also of interest on their own, and in other applications such as differential games, control of stochastic and jump processes, and robust control theory, etc.

Authors: Pierre Del Moral (Inria CQFD), Adrian Bishop and Sahani Pathiraja.

7.21. Probabilistic Safety Analysis of the Collision Between a Space Debris and a Satellite with an Island Particle Algorithm

Collision between satellites and space debris seldom happens, but the loss of a satellite by collision may have catastrophic consequences both for the satellite mission and for the space environment. To support the decision to trigger o a collision avoidance manoeuver, an adapted tool is the determination of the collision probability between debris and satellite. This probability estimation can be performed with rare event simulation techniques when Monte Carlo techniques are not enough accurate. In this chapter, we focus on analyzing the inuence of dierent simulation parameters (such as the drag coecient) that are set for to simplify the simulation, on the collision probability estimation. A bad estimation of these simulation parameters can strongly modify rare event probability estimations. We design here a new island particle Markov chain Monte

Carlo algorithm to determine the parameters that, in case of bad estimation, tend to increase the collision probability value. This algorithm also gives an estimate of the collision probability maximum taking into account the likelihood of the parameters. The principles of this statistical technique are described throughout this chapter.

Authors: Pierre Del Moral (Inria CQFD), Christelle Vergé, Jérôme Morio and Juan Carlos Dolado Pérez.

7.22. Biased online parameter inference for state-space models

We consider Bayesian online static parameter estimation for state-space models. This is a very important problem, but is very computationally challenging as the state-of-the art methods that are exact, often have a computational cost that grows with the time parameter; perhaps the most successful algorithm is that of SM C2 (Chopin et al., J R Stat Soc B 75: 397–426 2013). We present a version of the SM C2 algorithm which has computational cost that does not grow with the time parameter. In addition, under assumptions, the algorithm is shown to provide consistent estimates of expectations w.r.t. the posterior. However, the cost to achieve this consistency can be exponential in the dimension of the parameter space; if this exponential cost is avoided, typically the algorithm is biased. The bias is investigated from a theoretical perspective and, under assumptions, we find that the bias does not accumulate as the time parameter grows. The algorithm is implemented on several Bayesian statistical models.

Authors: Pierre Del Moral (Inria CQFD), Ajay Jasra and Yan Zhou.

7.23. Multilevel Sequential Monte Carlo Samplers for Normalizing Constants

This work considers the sequential Monte Carlo (SMC) approximation of ratios of normalizing constants associated to posterior distributions which in principle rely on continuum models. Therefore, the Monte Carlo estimation error and the discrete approximation error must be balanced. A multilevel strategy is utilized to substantially reduce the cost to obtain a given error level in the approximation as compared to standard estimators. Two estimators are considered and relative variance bounds are given. The theoretical results are numerically illustrated for the example of identifying a parametrized permeability in an elliptic equation given point-wise observations of the pressure.

Authors: Pierre Del Moral (Inria CQFD), Ajay Jasra, Kody Law and Yan Zhou.

7.24. Multilevel sequential Monte Carlo: Mean square error bounds under verifiable conditions

In this article, we consider the multilevel sequential Monte Carlo (MLSMC) method of Beskos et al. (Stoch. Proc. Appl. [to appear]). This is a technique designed to approximate expectations w.r.t. probability laws associated to a discretization. For instance, in the context of inverse problems, where one discretizes the solution of a partial differential equation. The MLSMC approach is especially useful when independent, coupled sampling is not possible. Beskos et al. show that for MLSMC the computational effort to achieve a given error, can be less than independent sampling. In this article we significantly weaken the assumptions of Beskos et al., extending the proofs to non-compact state-spaces. The assumptions are based upon multiplicative drift conditions as in Kontoyiannis and Meyn (Electron. J. Probab. 10 [2005]: 61–123). The assumptions are verified for an example.

Authors: Pierre Del Moral (Inria CQFD), Ajay Jasra and Kody Law.

7.25. Biased Online Parameter Inference for State-Space Models

We consider Bayesian online static parameter estimation for state-space models. This is a very important problem, but is very computationally challenging as the state-of-the art methods that are exact, often have a computational cost that grows with the time parameter; perhaps the most successful algorithm is that of SM C2 (Chopin et al., J R Stat Soc B 75: 397–426 2013). We present a version of the SM C2 algorithm which has computational cost that does not grow with the time parameter. In addition, under assumptions, the algorithm is shown to provide consistent estimates of expectations w.r.t. the posterior. However, the cost to achieve this consistency can be exponential in the dimension of the parameter space; if this exponential cost is avoided, typically the algorithm is biased. The bias is investigated from a theoretical perspective and, under assumptions, we find that the bias does not accumulate as the time parameter grows. The algorithm is implemented on several Bayesian statistical models.

Authors: Pierre Del Moral (Inria CQFD), Ajay Jasra and Yan Zhou.

7.26. Valuation of Barrier Options using Sequential Monte Carlo

Sequential Monte Carlo (SMC) methods have successfully been used in many applications in engineering, statistics and physics. However, these are seldom used in financial option pricing literature and practice. This paper presents SMC method for pricing barrier options with continuous and discrete monitoring of the barrier condition. Under the SMC method, simulated asset values rejected due to barrier condition are re-sampled from asset samples that do not breach the barrier condition improving the efficiency of the option price estimator; while under the standard Monte Carlo many simulated asset paths can be rejected by the barrier condition making it harder to estimate option price accurately. We compare SMC with the standard Monte Carlo method and demonstrate that the extra effort to implement SMC when compared with the standard Monte Carlo is very little while improvement in price estimate can be significant. Both methods result in unbiased estimators for the price converging to the true value as $1/\sqrt{M}$, where M is the number of simulations (asset paths). However, the variance of SMC estimator is smaller and does not grow with the number of time steps when compared to the standard Monte Carlo. In this paper we demonstrate that SMC can successfully be used for pricing barrier options. SMC can also be used for pricing other exotic options and also for cases with many underlying assets and additional stochastic factors such as stochastic volatility; we provide general formulas and references.

Authors: Pierre Del Moral (Inria CQFD) and Pavel V. Shevchenko.

8. Bilateral Contracts and Grants with Industry

8.1. Bilateral Contracts with Industry

8.1.1. DCNS

Participants: Huilong Zhang, Jonatha Anselmi, François Dufour, Dann Laneuville.

The increasing complexity of warfare submarine missions has led DCNS to study new tactical help functions for underwater combat management systems. In this context, the objective is to find optimal trajectories according to the current mission type by taking into account sensors, environment and surrounding targets. This problem has been modeled as a discrete-time Markov decision process with finite horizon. A quantization technique has been applied to discretize the problem in order to get a finite MDP for which standard methods such as the dynamic and/or the linear programming approaches can be applied. Different kind of scenarios have been considered and studied.

8.1.2. Thales Optronique

Participants: Benoîte de Saporta, François Dufour, Alizée Geeraert.

Maintenance, impulse control, failure, optimization. The objective of this grant in collaboration with Thales Optronique was to optimize the maintenance of a multi-component equipment that can break down randomly. The underlying problem was to choose the best dates to repair or replace components in order to minimize a cost criterion that takes into account costs of maintenance but also the cost associated to the unavailability of the system for the customer. This industrial process has been modeled by a piecewise deterministic Markov process (PDMP) and the maintenance problem has been formalized as an impulse control problem. We have applied an approximation method based on a quantization technique of the post jump location and inter-arrival time Markov chain naturally embedded in the PDMP, and a path-adapted time discretization grids to get an approximation of the value function. We have shown the existence of control strategies that can outperform reference control policies used by Thales Optronique. It remains to provide the explicit form of such strategies. This is actually the objective of a new collaboration with Thales Optronique that started in October 2017 funded by the Fondation Mathématique Jacques Hadamard.

8.1.3. Lyre: ADEQWAT project

Participants: François Dufour, Alexandre Genadot, Jérôme Saracco.

Stochastic modelling, Optimization. This project has just started in November 2017. The topic of this collaboration with Lyre, l'Agence de l'eau Adour-Garonne and ENSEGID is the modeling of the uncertainties in the Water demand adequacy in a context of global climate change. A PhD thesis (2018-2021) is part of this project.

9. Partnerships and Cooperations

9.1. National Initiatives

9.1.1. ANR Piece (2013-2017) of the program Jeunes chercheuses et jeunes chercheurs of the ANR

Piecewise Deterministic Markov Processes (PDMP) are non-diffusive stochastic processes which naturally appear in many areas of applications as communication networks, neuron activities, biological populations or reliability ofcomplex systems. Their mathematical study has been intensively carried out in the past two decades but many challenging problems remain completely open. This project aims at federating a group of experts with different backgrounds (probability, statistics, analysis, partial derivative equations, modelling) in order to pool everyone's knowledge and create new tools to study PDMPs. The main lines of the project relate to estimation, simulation and asymptotic behaviors (long time, large populations, multi-scale problems) in the various contexts of application.

9.1.2. ANR StocMC (2014-2018) of the program Project Blanc of the ANR

The involved research groups are Inria Rennes/IRISA Team SUMO; Inria Rocquencourt Team Lifeware; LIAFA University Paris 7; Bordeaux University.
The aim of this research project is to develop scalable model checking techniques that can handle large stochastic systems. Large stochastic systems arise naturally in many different contexts, from network systems to system biology. A key stochastic model we will consider is from the biological pathway of apoptosis, the programmed cell death.

9.1.3. ANR BNPSI: Bayesian Non Parametric methods for Signal and Image Processing

Statistical methods have become more and more popular in signal and image processing over the past decades. These methods have been able to tackle various applications such as speech recognition, object tracking, image segmentation or restoration, classification, clustering, etc. We propose here to investigate the use of Bayesian nonparametric methods in statistical signal and image processing. Similarly to Bayesian parametric methods, this set of methods is concerned with the elicitation of prior and computation of posterior distributions, but now on infinite-dimensional parameter spaces. Although these methods have become very popular in statistics and machine learning over the last 15 years, their potential is largely underexploited in signal and image processing. The aim of the overall project, which gathers researchers in applied probabilities, statistics, machine learning and signal and image processing, is to develop a new framework for the statistical signal and image processing communities. Based on results from statistics and machine learning we aim at defining new models, methods and algorithms for statistical signal and image processing. Applications to hyperspectral image analysis, image segmentation, GPS localization, image restoration or space-time tomographic reconstruction will allow various concrete illustrations of the theoretical advances and validation on real data coming from realistic contexts.

9.1.4. Gaspard Monge Program for Optimisation and Operational Research (2017-2018)

The involved research groups are Inria Bordeaux Sud-Ouest Team CQFD and Thales Optronique. This new collaboration with Thales Optronique that started in October 2017 is funded by the Fondation Mathématique Jacques Hadamard. This is the continuation of the PhD Thesis of A. Geeraert. The objective of this project is to optimize the maintenance of a multi-component equipment that can break down randomly. The underlying problem is to choose the best dates to repair or replace components in order to minimize a cost criterion that takes into account costs of maintenance but also the cost associated to the unavailability of the system for the customer. In the PhD thesis of A. Geeraert, the model under consideration was rather simple and only a numerical approximation of the value function was provided. Here, our objective is more ambitious. A more realistic model will be considered and our aim is to provide a tractable quasi-optimal control strategy that can be applied in practice to optimize the maintenance of such equipments.

9.2. European Initiatives

9.2.1. Collaborations in European Programs, Except FP7 & H2020

- Program: Direcion General de Investigacion Cientifica y Tecnica, Gobierno de Espana
- Project acronym: GAMECONAPX
- Project title: Numerical approximations for Markov decision processes and Markov games
- Duration: 01/2017 - 12/2019
- Coordinator: Tomas Prieto-Rumeau, Department of Statistics and Operations Research, UNED (Spain)
- Abstract:

  This project is funded by the Gobierno de Espana, Direcion General de Investigacion Cientifica y Tecnica (reference number: MTM2016-75497-P) for three years to support the scientific collaboration between Tomas Prieto-Rumeau, Jonatha Anselmi and Francois Dufour. This research project is concerned with numerical approximations for Markov decision processes and Markov games. Our goal is to propose techniques allowing to approximate numerically the optimal value function and the optimal strategies of such problems. Although such decision models have been widely studied theoretically and, in general, it is well known how to characterize their optimal value function and their optimal strategies, the explicit calculation of these optimal solutions is not possible except for a few particular cases. This shows the need for numerical procedures to estimate or to approximate the optimal solutions of Markov decision processes and Markov games, so that the decision maker can really have at hand some approximation of his optimal strategies and his optimal value function. This project will explore areas of research that have been, so far, very little investigated. In this sense, we expect our techniques to be a breakthrough in the field of numerical methods for continuous-time Markov decision processes, but particularly in the area of numerical methods for Markov game models. Our techniques herein will cover a wide range of models, including discrete- and continuous-time models, problems with unbounded cost and transition rates, even allowing for discontinuities of these rate functions. Our research results will combine, on one hand, mathematical rigor (with the application of advanced tools from probability and measure theory) and, on the other hand, computational efficiency (providing accurate and ?applicable? numerical methods). In this sense, particular attention will be paid to models of practical interest, including population dynamics, queueing systems, or birth-and-death processes, among others. So, we expect to develop a generic and robust methodology in which, by suitably specifying the data of the decision problem, an algorithm will provide the approximations of the value function and the optimal strategies. Therefore, the results that we intend to obtain in this research project will be of interest for researchers in the fields of Markov decision processes and Markov games, both for the theoretical and the applied or practitioners communities

9.3. International Initiatives

9.3.1. Inria International Partners

9.3.1.1. Declared Inria International Partners

Tree-Lab, ITT. TREE-LAB is part of the Cybernetics research line within the Engineering Science graduate program offered by the Department of Electric and Electronic Engineering at Tijuana's Institute of Technology (ITT), in Tijuana Mexico. TREE-LAB is mainly focused on scientific and engineering research within the intersection of broad scientific fields, particularly Computer Science, Heuristic Optimization and Pattern Analysis. In particular, specific domains studied at TREE-LAB include Genetic Programming, Classification, Feature Based Recognition, Bio-Medical signal analysis and Behavior-Based Robotics. Currently, TREE-LAB incorporates the collaboration of several top researchers, as well as the participation of graduate (doctoral and masters) and undergraduate students, from ITT. Moreover, TREE-LAB is actively collaborating with top researchers from around the world, including Mexico, France, Spain, Portugal and USA.

9.4. International Research Visitors

9.4.1. Visits of International Scientists

Tomas Prieto-Rumeau (Department of Statistics and Operations Research, UNED, Madrid, Spain) visited the team during 2 weeks in 2017. The main subject of the collaboration is the approximation of Markov Decision Processes.

Oswaldo Costa (Escola Politécnica da Universidade de São Paulo, Brazil) collaborate with the team on the theoretical aspects of continuous control of piecewise-deterministic Markov processes. He visited the team during two weeks in 2017.

10. Dissemination

10.1. Promoting Scientific Activities

10.1.1. Scientific Events Organisation

10.1.1.1. Member of the Organizing Committees

P. Legrand was co-organizer of EA 2017.

F. Dufour has been a member of the Organizing Committee of the SIAM Conference on Control and Its Applications (CT17) in Pittsburgh, USA, 2017.

10.1.2. Scientific Events Selection

10.1.2.1. Chair of Conference Program Committees

P. Legrand was chair for EA 2017.

10.1.2.2. Member of the Conference Program Committees

J. Anselmi has been a member of the TPC of the following international conferences: VALUETOOLS-2017, ASMTA-2017 and IFIP Performance 2017.

P. Legrand has been a member of the PC of the following international conferences: EA 2017.

10.1.3. Journal

10.1.3.1. Member of the Editorial Boards

P. Del Moral is an associate editor for the journal Stochastic Analysis and Applications since 2001.

P. Del Moral is an associate editor for the journal Revista de Matematica: Teoria y aplicaciones since 2009.

P. Del Moral is an associate editor for the journal Applied Mathematics and Optimization since 2009.

F. Dufour is associate editor of the journal: SIAM Journal of Control and Optimization since 2009.

F. Dufour is the representative of the SIAM activity group in control and system theory for the journal SIAM News since 2014.

J. Saracco is an associate editor of the journal Case Studies in Business, Industry and Government Statistics (CSBIGS) since 2006.

10.1.3.2. Reviewer - Reviewing Activities

All the members of CQFD are regular reviewers for several international journals and conferences in applied probability, statistics and operations research.

10.1.4. Invited Talks

Pierrick Legrand was invited to give a talk on IA in Thales center in october.

Pierre del Moral gave several invited talks listed at the following address: http://people.bordeaux.inria.fr/pierre.delmoral/conf.html

10.1.5. Leadership within the Scientific Community

P. Legrand was the scientific leader of the programs HUMO 3 (with UBX and IMS) and MICRO-DOPPLER (with Thales and IMS) in the context of the GIS ALBATROS.

10.1.6. Scientific Expertise

Pierrick Legrand has been reviewer for the ANR generics projects in 2017.

J. Saracco is elected member of the council of the *Société Française de Statistique* (SFdS, French Statistical Society).

J. Saracco was vice president of SFdS from 2014 to 2016.

10.1.7. Research Administration

J. Saracco is deputy director of IMB (Institut de Mathématiques de Bordeaux, UMR CNRS 5251) since 2015.

M. Chavent is member of the national evaluation committee of Inria.

M. Chavent and Pierrick Legrand are members of the council of the Institut de Mathématique de Bordeaux.

Pierrick Legrand was the director of the Ressources Center Victoire of the UF Mathematics and interactions until may 2017.

10.2. Teaching - Supervision - Juries

10.2.1. Teaching

- Licence : J. Anselmi, Probabilités et statistiques, 20 heures, L3, Institut Polytechnique de Bordeaux, école ENSEIRB-MATMECA, filiÃšre Télécommunications, France.
- Licence : J. Anselmi, Probabilités et statistiques, 16 heures, L3, Institut Polytechnique de Bordeaux, école ENSEIRB-MATMECA, filiÃšre Electronique, France.
- Licence : J. Anselmi, Probabilités et statistiques, 48 heures, niveau L3, Institut Polytechnique de Bordeaux, école ENSEIRB-MATMECA, filiÃšre Mathématique et Mécanique, France.
- Licence: M. Chavent, Analyse des données, 15 ETD, L3, Bordeaux university, France
- License: M. Chavent, Modélisation statistique, 15 ETD, niveau L3, Bordeaux university, France
- Master : M. Chavent, Apprentissage automatique, 50 ETD, niveau M2, Bordeaux university, France
- Licence : F. Dufour, Probabilités et statistiques, 70h, first year of école ENSEIRB-MATMECA, Institut Polytechnique de Bordeaux, France.
- Master : F. Dufour, Méthodes numériques pour la fiabilité, 36h, third year of école ENSEIRB-MATMECA, Institut Polytechnique de Bordeaux, France.
- P. Legrand, AlgÃšbre (responsable de l'UE), Licence 1 SCIMS (108 heures)
- P. Legrand, Informatique pour les mathÃÂ©matiques (responsable de l'UE), Licence 1 et Licence 2 (36 heures)
- P. Legrand, Espaces Euclidiens. (responsable de l'UE), Licence 2 SCIMS (54 heures)
- Licence: J. Saracco, Probability and Descriptive statistics, 27h, L3, First year of ENSC - Bordeaux INP, France
- Licence: J. Saracco, Mathematical statistics, 20h, L3, First year of ENSC - Bordeaux INP, France
- Licence: J. Saracco, Data analysis (multidimensional statistics), 20h, L3, First year of ENSC - Bordeaux INP, France
- Master: J. Saracco, Statistical modeling, 27h, M1, Second year of ENSC - Bordeaux INP, France
- Master: J. Saracco, Applied probability and Statistics, 40h, M1, Second year of ENSCBP - Bordeaux INP, France
- Master: J. Saracco, Probability and Statistics, 12h, M2, Science Po Bordeaux, France
- A. Genadot, Probabilités de bases (18h), Licence MIASHS premiÃšre année, Université de Bordeaux.
- A. Genadot, Statistiques de bases (18h), Licence MIASHS premiÃšre année, Université de Bordeaux.
- A. Genadot, Probabilités (36h), Licence MIASHS deuxiÃšme année, Université de Bordeaux.
- A. Genadot, Processus (18h), Licence MIASHS troisiÃšme année, Université de Bordeaux.
- A. Genadot, Modélisation statistique (18h), Licence MIASHS troisiÃšme année, Université de Bordeaux.
- A. Genadot, Martingales (25h), Master MIMSE premiÃšre année, Université de Bordeaux.
- A. Genadot, Probabilités (20h), Master MEEF premiÃšre année, Université de Bordeaux.

10.2.2. Supervision

- PhD completed : Alizé Geeraert, Contrôle optimal des processus Markoviens déterministes par morceaux et application à la maintenance, University of Bordeaux, supervised by B. de Saporta and F. Dufour (defense in June 2017).

- PhD in progress : Ines Jlassi, Contributions à la régression inverse par tranches et à l'estimation non para métrique des quantiles conditionnels, University of Monastir (Tunisia), September 2013, supervised by J. Saracco and L. Ben Abdelghani Bouraoui.
- PhD in progress : Hadrien Lorenzo, Analyses de données longitudinales de grandes dimensions appliquées aux essais vaccinaux contre le VIH et Ebola, University of Bordeaux, September 2016, supervised by J. Saracco and R. Thiebaut.
- PhD in progress : Tiffany Cherchi, "Automated optimal fleet management policy for airborne equipment", Montpellier University, since 2017, supervised by B. De Saporta and F. Dufour.
- PhD in progress : Chloé Pasin, "Modelisation et optimisation de la réponse vaccinale. Application au VIH et Ebola", Bordeaux University, since 2015, supervised by F. Dufour and R. Thiebaut.
- PhD in progress : Maud Joubaud, "Branching piecewise deterministic Markov processes, applications to cell biology", Montpellier University, since 2016, supervised by B. De Saporta and B. Cloez.
- PhD completed : Emigdio Z. Flores, Human mental states classification using EEG by means of Genetic Programming, ITT Tijuana, supervised by L. Trujillo and P. Legrand (defense in July 2017).

10.2.3. Juries

J. Saracco is vice president of the french statistical society (SFdS).

11. Bibliography

Publications of the year

Doctoral Dissertations and Habilitation Theses

[1] A. GEERAERT. *Stochastic optimal control for piecewise deterministic Markov processes and application to maintenance optimization*, Université de Bordeaux, June 2017, https://tel.archives-ouvertes.fr/tel-01557969

[2] E. Z-FLORES. *Human mental states classification using EEG by means of Genetic Programming*, ITT, Instituto tecnologico de Tijuana, July 2017, https://hal.inria.fr/tel-01668672

Articles in International Peer-Reviewed Journals

[3] J. ANSELMI. *Asymptotically optimal open-loop load balancing*, in "Queueing Systems", September 2017, pp. 1–23 [*DOI :* 10.1007/s11134-017-9547-9], https://hal.inria.fr/hal-01614892

[4] J. ANSELMI, D. ARDAGNA, J. C. S. LUI, A. WIERMAN, Y. XU, Z. YANG. *The Economics of the Cloud: Price Competition and Congestion*, in "ACM Transactions on Modeling and Performance Evaluation of Computing Systems", September 2017, vol. 2, n^o 4, pp. 1 - 23 [*DOI :* 10.1145/3086574], https://hal.inria.fr/hal-01615172

[5] R. AZAÏS, A. GENADOT. *A new characterization of the jump rate for piecewise-deterministic Markov processes with discrete transitions*, in "Communication in Statistics - Theory and Methods", 2017, https://arxiv.org/abs/1606.06130 , forthcoming, https://hal.archives-ouvertes.fr/hal-01334847

[6] A. N. BISHOP, P. DEL MORAL. *On the Stability of Kalman–Bucy Diffusion Processes*, in "SIAM Journal on Control and Optimization", November 2017, vol. 55, n^o 6, pp. 4015 - 4047 [*DOI :* 10.1137/16M1102707], https://hal.inria.fr/hal-01669244

[7] A. N. BISHOP, P. DEL MORAL, S. D. PATHIRAJA. *Perturbations and projections of Kalman–Bucy semigroups*, in "Stochastic Processes and their Applications", November 2017, pp. 1-48 [*DOI :* 10.1016/J.SPA.2017.10.006], https://hal.inria.fr/hal-01669084

[8] M. CHAVENT, V. KUENTZ-SIMONET, A. LABENNE, J. SARACCO. *ClustGeo: an R package for hierarchical clustering with spatial constraints*, in "Computational Statistics", 2017, pp. 1-26, https://arxiv.org/abs/1707.03897 , forthcoming, https://hal.archives-ouvertes.fr/hal-01664018

[9] E. COSTA, B. DE SAPORTA. *Linear minimum mean square filters for Markov jump linear systems*, in "IEEE Transactions on Automatic Control", 2017, vol. 62, n^{o} 7, pp. 3567-3572, https://arxiv.org/abs/1601.00772 [*DOI :* 10.1109/TAC.2017.2692180], https://hal.archives-ouvertes.fr/hal-01251334

[10] O. L. D. V. COSTA, F. DUFOUR. *Zero-Sum Discounted Reward Criterion Games for Piecewise Deterministic Markov Processes*, in "Applied Mathematics and Optimization", 2017, forthcoming [*DOI :* 10.1007/s00245-017-9416-2], https://hal.archives-ouvertes.fr/hal-01643056

[11] B. DE SAPORTA, F. DUFOUR, A. GEERAERT. *Optimal strategies for impulse control of piecewise deterministic Markov processes*, in "Automatica", 2017, vol. 77, pp. 219-229, https://arxiv.org/abs/1603.07859 [*DOI :* 10.1016/J.AUTOMATICA.2016.11.039], https://hal.archives-ouvertes.fr/hal-01294286

[12] B. DE SAPORTA, F. DUFOUR, C. NIVOT. *Partially observed optimal stopping problem for discrete-time Markov processes*, in "4OR: A Quarterly Journal of Operations Research", 2017, https://arxiv.org/abs/1602.04609 [*DOI :* 10.1007/S10288-016-0337-8], https://hal.archives-ouvertes.fr/hal-01274645

[13] P. DEL MORAL, A. JASRA. *A sharp first order analysis of Feynman–Kac particle models, Part I: Propagation of chaos*, in "Stochastic Processes and their Applications", January 2018, vol. 128, n^{o} 1, pp. 332 - 353 [*DOI :* 10.1016/J.SPA.2017.04.007], https://hal.inria.fr/hal-01669381

[14] P. DEL MORAL, A. JASRA. *A sharp first order analysis of Feynman–Kac particle models, Part II: Particle Gibbs samplers*, in "Stochastic Processes and their Applications", January 2018, vol. 128, n^{o} 1, pp. 354 - 371 [*DOI :* 10.1016/J.SPA.2017.05.001], https://hal.inria.fr/hal-01669385

[15] P. DEL MORAL, A. KURTZMANN, J. TUGAUT. *On the Stability and the Uniform Propagation of Chaos of a Class of Extended Ensemble Kalman–Bucy Filters*, in "SIAM Journal on Control and Optimization", 2017, vol. 55, n^{o} 1, pp. 119 - 155, https://arxiv.org/abs/1606.08256 [*DOI :* 10.1137/16M1087497], https://hal.inria.fr/hal-01469198

[16] P. DEL MORAL, D. VILLEMONAIS. *Exponential mixing properties for time inhomogeneous diffusion processes with killing*, in "Bernoulli", May 2018, vol. 24, n^{o} 2, pp. 1010 - 1032, https://arxiv.org/abs/1412.2627 [*DOI :* 10.3150/16-BEJ845], https://hal.inria.fr/hal-01593882

[17] A. GENADOT. *Averaging for some simple constrained Markov processes*, in "Probability and Mathematical Statistics", 2017, https://arxiv.org/abs/1606.07577 , forthcoming, https://hal.archives-ouvertes.fr/hal-01336752

[18] F. GIRAUD, P. DEL MORAL. *Nonasymptotic analysis of adaptive and annealed Feynman–Kac particle models*, in "Bernoulli", 2017, vol. 23, n^{o} 1, pp. 670 - 709, https://arxiv.org/abs/1209.5654 [*DOI :* 10.3150/14-BEJ680], https://hal.inria.fr/hal-01469193

[19] I. JLASSI, J. SARACCO. *Variable importance assessment in sliced inverse regression for variable selection*, in "Communications in Statistics - Simulation and Computation", September 2017, pp. 1 - 31 [*DOI :* 10.1080/03610918.2017.1375522], https://hal.archives-ouvertes.fr/hal-01662513

[20] Y. MARTINEZ, E. NAREDO, L. TRUJILLO, P. LEGRAND, U. LOPEZ. *A comparison of fitness-case sampling methods for genetic programming*, in "Journal of Experimental and Theoretical Artificial Intelligence", 2017, https://hal.inria.fr/hal-01389047

[21] P. D. MORAL, A. JASRA, K. J. H. LAW, Y. ZHOU. *Multilevel Sequential Monte Carlo Samplers for Normalizing Constants*, in "ACM Transactions on Modeling and Computer Simulation", September 2017, vol. 27, n^{o} 3, pp. 1 - 22 [*DOI :* 10.1145/3092841], https://hal.inria.fr/hal-01669115

[22] P. D. MORAL, A. JASRA, Y. ZHOU. *Biased Online Parameter Inference for State-Space Models*, in "Methodology and Computing in Applied Probability", September 2017, vol. 19, n^{o} 3, pp. 727 - 749 [*DOI :* 10.1007/s11009-016-9511-x], https://hal.inria.fr/hal-01669132

[23] H. ZHANG, B. DE SAPORTA, F. DUFOUR, D. LANEUVILLE, A. NÈGRE. *Stochastic Control of Observer Trajectories in Passive Tracking with Acoustic Signal Propagation Optimization*, in "IET Radar Sonar and Navigation", 2017, forthcoming, https://hal.inria.fr/hal-01610189

Articles in National Peer-Reviewed Journals

[24] M. P. ELLIES-OURY, G. CANTALAPIEDRA HIJAR, D. DURAND, D. GRUFFAT, A. LISTRAT, D. D. MICOL, I. ORTIGUES-MARTY, J.-F. HOCQUETTE, M. CHAVENT, J. SARACCO, B. PICARD. *Une nouvelle approche méthodologique pour piloter la conduite en élevage*, in "Viandes & Produits Carnés", 2017, vol. 33, n^{o} 3, pp. 1-8, https://hal.archives-ouvertes.fr/hal-01664209

Invited Conferences

[25] A. CLEYNEN, B. DE SAPORTA. *Optimal Stopping for Change-point Detection of Piecewise Deterministic Markov Processes*, in "CT 2017 - SIAM Conference on Control and Its Applications", Pittsburgh, United States, 2017, https://hal.archives-ouvertes.fr/hal-01563234

[26] E. COSTA, B. DE SAPORTA. *Clustered Information Filter for Markov Jump Linear Systems*, in "CT 2017 - SIAM Conference on Control and Its Applications", Pittsburgh, United States, July 2017, https://hal.archives-ouvertes.fr/hal-01563236

[27] B. DE SAPORTA, F. DUFOUR. *Méthode numérique pour le contrôle des PDMP*, in "Statistique pour les processus de Markov déterministes par morceaux", Nancy, France, Romain Azaïs, Florian Bouguet, February 2017, https://hal.archives-ouvertes.fr/hal-01549078

[28] F. DUFOUR, A. GENADOT, A. PIUNOVSKIY. *The Expected Total Reward Criterion for Markov Decision Processes*, in "IMA and OR Society Conference on Mathematics of Operational Research", Birmingham, United Kingdom, April 2017, https://hal.inria.fr/hal-01665757

[29] F. DUFOUR, T. PRIETO-RUMEAU. *Approximations for Average Markov Decision Processes in Continuous-time*, in "SIAM Conference on Control and its Applications", Pittsburgh, United States, July 2017, https://hal.inria.fr/hal-01665755

International Conferences with Proceedings

[30] H. ZHANG, F. DUFOUR, J. ANSELMI, D. LANEUVILLE, A. NÈGRE. *Piecewise Optimal Trajectories of Observer for Bearings-Only Tracking by Quantization*, in "20th International Conference on Information Fusion", Xi'an, China, July 2017, pp. 1458-1464, https://hal.archives-ouvertes.fr/hal-01591013

[31] H. ZHANG, F. DUFOUR, J. ANSELMI, D. LANEUVILLE, A. NÈGRE. *Piecewise Optimal Trajectories of Observer for Bearings-Only Tracking of Maneuvering Target*, in "Proceedings of IEEE Aerospace Conference", Bigsky, United States, March 2018, https://hal.inria.fr/hal-01649429

Conferences without Proceedings

[32] S. BADILLO, M. CHAVENT, M. JOLIOT, N. TZOURIO-MAZOYER, J. SARACCO, B. MAZOYER. *Are production, reading, and listening co-lateralized ? fMRI investigation in 296 healthy individuals*, in "OHBM 2017 - Organization for Human Brain Mapping", Vancouver, Canada, June 2017, https://hal.archives-ouvertes.fr/hal-01663987

[33] M. CHAVENT, G. CHAVENT. *Approche bloc en ACP group-sparse: le package sparsePCA*, in "AngletR 2017 - 6ème Rencontres R", Anglet, France, June 2017, 1 p. , https://hal.archives-ouvertes.fr/hal-01662652

[34] H. LORENZO, R. THIEBAUT, J. SARACCO. *Multi-block high-dimensional lasso-penalized analysis with imputation of missing data applied to postgenomic data in an Ebola vaccine trial*, in "Annual workshop on Statistical Methods for Post Genomic Data - SMPGD 2018", Montpellier, France, January 2018, https://hal.archives-ouvertes.fr/hal-01664610

[35] J. SARACCO, I. JLASSI. *Evaluation de l'importance des variables dans la méthode SIR (Sliced Inverse Regression)*, in "JDS 2017 - 49èmes Journées de Statistique", Avignon, France, May 2017, https://hal.archives-ouvertes.fr/hal-01664081

Scientific Books (or Scientific Book chapters)

[36] T. LEONARDO, E. Z-FLORES, P. S. JUAREZ SMITH, P. LEGRAND, S. SILVA, M. CASTELLI, L. VANNESCHI, O. SCHÜTZE, L. MUNOZ. *Local Search is Underused in Genetic Programming*, in "Genetic Programming Theory and Practice XIV", A. ARBOR (editor), Springer, 2017, https://hal.inria.fr/hal-01388426

Research Reports

[37] A. N. BISHOP, P. DEL MORAL. *Matrix product moments in normal variables*, Arxiv, May 2017, 41 p. , https://arxiv.org/abs/1703.00353 , https://hal.inria.fr/hal-01593836

[38] A. N. BISHOP, P. DEL MORAL. *On the stability of Kalman-Bucy diffusion processes*, Arxiv, August 2017, https://arxiv.org/abs/1610.04686 , https://hal.inria.fr/hal-01593875

[39] A. N. BISHOP, P. DEL MORAL, A. NICLAS. *A perturbation analysis of stochastic matrix Riccati diffusions*, Arxiv, September 2017, https://arxiv.org/abs/1709.05071 , https://hal.inria.fr/hal-01593830

[40] A. N. BISHOP, P. DEL MORAL, S. D. PATHIRAJA. *Perturbations and Projections of Kalman-Bucy Semigroups Motivated by Methods in Data Assimilation*, Arxiv, January 2017, https://arxiv.org/abs/1701.05798 , https://hal.inria.fr/hal-01593872

[41] P. DEL MORAL, A. NICLAS. *A Taylor expansion of the square root matrix functional*, Arxiv, May 2017, https://arxiv.org/abs/1705.08561 , https://hal.inria.fr/hal-01593833

Other Publications

[42] R. AZAÏS, A. GENADOT. *Estimation of the average number of continuous crossings for non-stationary non-diffusion processes*, November 2017, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01499914

[43] R. AZAÏS, A. GENADOT, B. HENRY. *Inference for conditioned Galton-Watson trees from their Harris path*, June 2017, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01360650

[44] A. N. BISHOP, P. DEL MORAL, K. KAMATANI, B. RÉMILLARD. *On one-dimensional Riccati diffusions*, December 2017, working paper or preprint, https://hal.inria.fr/hal-01662597

[45] A. N. BISHOP, P. DEL MORAL, A. NICLAS. *An introduction to Wishart matrix moments*, December 2017, working paper or preprint [*DOI :* 10.10864], https://hal.inria.fr/hal-01662575

[46] M. CHAVENT, G. CHAVENT. *Group-sparse block PCA and explained variance*, December 2017, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01662605

[47] M. CHAVENT, V. KUENTZ-SIMONET, A. LABENNE, J. SARACCO. *Multivariate Analysis of Mixed Data: The R Package PCAmixdata*, December 2017, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01662595

[48] B. CLOEZ, R. DESSALLES, A. GENADOT, F. MALRIEU, A. MARGUET, R. YVINEC. *Probabilistic and Piecewise Deterministic models in Biology*, June 2017, https://arxiv.org/abs/1706.09163 - Proceedings of the journées MAS 2016, session Probabilistic and Piecewise Deterministic models in Biology, https://hal.archives-ouvertes.fr/hal-01549687

[49] B. DE SAPORTA, A. CLEYNEN. *Change-point detection for Piecewise Deterministic Markov Processes*, September 2017, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01596670

[50] I. JLASSI, J. SARACCO. *A Smooth Nonparametric Estimator of a Conditional Quantile*, 2017, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01664075

References in notes

[51] N. DUAN, K.-C. LI. *Slicing regression: a link-free regression method*, in "Ann. Statist.", 1991, vol. 19, n^o 2, pp. 505–530, http://dx.doi.org/10.1214/aos/1176348109

[52] R. DUDA, P. HART, D. STORK. *Pattern Classification*, John Wiley, 2001

[53] K.-C. LI. *Sliced inverse regression for dimension reduction*, in "J. Amer. Statist. Assoc.", 1991, vol. 86, n^o 414, pp. 316–342, With discussion and a rejoinder by the author