IN PARTNERSHIP WITH:
CNRS

Université de Bordeaux

Activity Report 2013

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

Keywords: Stochastic Methods, Statistical Methods, Control Theory, Optimization, Data Analysis, Reliability

Creation of the Project-Team: 2009 January 01.

1. Members

Faculty Members

François Dufour [Team leader, IPB, Professor, HdR]
Marie Chavent [Univ. Bordeaux II, Associate Professor, HdR]
Benoîte de Saporta [Univ. Bordeaux IV, Associate Professor, HdR]
Anne Gégout-Petit [Univ. Nancy I, Associate Professor, until Aug 2013, HdR]
Jérôme Saracco [IPB, Professor, HdR]
Huilong Zhang [Univ. Bordeaux I, Associate Professor]

PhD Students

Camille Baysse [Thales, granted by CIFRE, until Nov 2013]
Isabelle Charlier [Univ. Bordeaux and Univ. Bruxelles]
Raphaël Coudret [Univ. Bordeaux I, until Sep 2013]
Amaury Labenne [IRSTEA]
Shuxian Li [Univ. Bordeaux I]
Christophe Nivot [Inria, from Mar 2013]
Laurent Vézard [CNRS, until Oct 2013]
Romain Azaïs [ANR, until Oct 2013]

Visiting Scientists

Eduardo Costa [Univ. São Paulo, from Jul 2013 until Aug 2013]
Oswaldo Luiz Do Valle Costa [Univ. São Paulo, from Sept. 1, until Sept. 14, 2013]
Alexei Piunovskiy [Univ. Liverpool, from June 10, until July 25, 2013]
Thomas Prieto Rumeau [UNED, spain, from May 20, until Mai 31, 2013]

Administrative Assistants

Catherine Cattaert Megrat [Inria, from Jan 2013]
Nicolas Jahier [Inria]

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 mathematical representations which are in essence dynamic and stochastic with discrete and/or continuous variables. This increasing 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 and various probabilistic phenomena;
- Estimation: developing theoretical and computational procedures in order to estimate and evaluate the parameters and the performance of the system;
- Optimization: developing theoretical and numerical control tools to optimize the performance and/or to maintain the system function in operating state.

2.2. Highlights of the Year

The paper *Optimal stopping for predictive maintenance of a structure subject to corrosion* [7] was nominated winner of the 2012 SAGE Best Paper Award by the editorial board of Journal of Risk 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. For example, in order to optimize the predictive maintenance of a system, it is necessary to choose the adequate model for its representation. The step of modeling is crucial before any estimation or computation of quantities related to its optimization. For this we have to represent all the different regimes of the system and the behavior of the physical variables under each of these regimes. Moreover, we must also select the dynamic variables which have a potential effect on the physical variable and the quantities of interest. The team CQFD works on the theory of Piecewise Deterministic Markov Processes (PDMP's) and on Markov Decision Processes (MDP's). These two classes of systems form general families of controlled stochastic processes suitable for the modeling of sequential decision-making problems in the continuous-time (PDMPs) and discrete-time (MDP's) context. They appear in many fields such as engineering, computer science, economics, operations research and constitute powerful class of processes for the modeling of complex system.
- 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. However, 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 [52] and Duan and Li [50]
  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 [51]
- Stochastic optimal control: optimal stopping, impulse control, continuous control, linear programming, singular perturbation, martingale problem.
  The first objective is to focus on the development of computational methods.
  - In the continuous-time context, stochastic control theory has from the numerical point of view, been mainly concerned with Stochastic Differential Equations (SDEs in short). From the practical and theoretical point of view, the numerical developments for this class of processes are extensive and largely complete. It capitalizes on the connection between SDEs and second order partial differential equations (PDEs in short) and the fact that the properties of the latter equations are very well understood. It is, however, hard to deny that the development of computational methods for the control of PDMPs has received little attention. One of the main reasons is that the role played by the familiar PDEs in the diffusion models is here played by certain systems of integro-differential equations for which there is not (and cannot be) a unified theory such as for PDEs as emphasized by M.H.A. Davis in his book. To the best knowledge of the team, there is only one attempt to tackle this difficult problem by O.L.V. Costa and M.H.A. Davis. The originality of our project consists in studying this unexplored area. It is very important to stress the fact that these numerical developments will give rise to a lot of theoretical issues such as type of approximations, convergence results, rates of convergence,....
  - Theory for MDP's has reached a rather high degree of maturity, although the classical tools such as value iteration, policy iteration and linear programming, and their various extensions, are not applicable in practice. We believe that the theoretical progress of MDP's must be in parallel with the corresponding numerical developments. Therefore, solving

MDP's numerically is an awkward and important problem both from the theoretical and practical point of view. In order to meet this challenge, the fields of neural networks, neuro-dynamic programming and approximate dynamic programming became recently an active area of research. Such methods found their roots in heuristic approaches, but theoretical results for convergence results are mainly obtained in the context of finite MDP's. Hence, an ambitious challenge is to investigate such numerical problems but for models with general state and action spaces. Our motivation is to develop theoretically consistent computational approaches for approximating optimal value functions and finding optimal policies.

Analysis of various problems arising in MDPs leads to a large variety of interesting mathematical problems. The second objective of the team is to study some theoretical aspects related to MDPs such as convex analytical methods and singular perturbation.

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 which gathers all the topics of team is a collaboration started in May 2010 with Thales Optronique on the subject of *optimization of the maintenance of a digital camera equipped with HUMS* (Health Unit Monitoring Systems). This subject is very interesting for us because it combines many aspects of our project. Classification tools will be used to select significant variables as the first step in the modeling of a digital camera. The model will then be analysed and estimated in order to optimize the maintenance.

A second example concerns the optimization of the maintenance date for an aluminum metallic structure subject to corrosion. It is a structure of strategic ballistic missile that is stored in a nuclear submarine missile launcher in peace-time and inspected with a given periodicity. The requirement for security on this structure is very strong. The mechanical stress exerted on the structure depends on its thickness. It is thus crucial to control the evolution of the thickness of the structure over time, and to intervene before the break.

A third example is the minimization of the acoustic signature of a submarine. The submarine has to chose its trajectory in order to minimize at each time step its observability by a surface ship following an unknown random trajectory.

However the spectrum of applications of the topics of the team is larger and may 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. Software and Platforms

5.1. Package edrGraphicalTools

This R package gives graphical tools for selecting the number of slices and the dimension of the model in SIR and SAVE approaches. It also provides the estimation of the reduction dimension subspace and the non parametric estimation of the link function using smoothing techniques. The package is available via the link http://cran.r-project.org/web/packages/edrGraphicalTools/index.html.

5.2. Package ClustOfVar

This R package is dedicated to cluster analysis of a set of variables. Variables can be quantitative, qualitative or a mixture of both. A new version 0.8 of the package is available since december 2013 via the link http://cran.r-project.org/web/packages/ClustOfVar/index.html. This version provides now a function to predict values of new observations on the synthetic variables of the clusters. This new function was used for supervised classification and variable selection in gene expressions data [42].

5.3. Package PCAmixdata

This package is dedicated to factorial analysis and rotation of quantitative data, qualitative data, or mixed data. The PCAMIX method, proposed in this package includes the ordinary principal component analysis (PCA) and multiple correspondence analysis (MCA) as special cases. Orthogonal varimax rotation of the principal components of PCAMIX is also implemented in this package.This year, a new method has been developped for Multiple Factorial Analysis in case of mixtures of quantitative and qualitative variables within groups. It was implemented in the package and presented to the 2èmes Rencontres R in Lyon [44], and to the 45èmes Journées de Statistique in Toulouse [41].

6. New Results

6.1. Nonparametric estimation of the jump rate for non-homogeneous marked renewal processes.

Participants: Romain Azais, François Dufour, Anne Gégout-Petit.

Non-homogeneous marked renewal process, nonparametric estimation, jump rate estimation, Nelson-Aalen estimator, asymptotic consistency, ergodicity of Markov chains

This work is devoted to the nonparametric estimation of the jump rate and the cumulative rate for a general class of non-homogeneous marked renewal processes, defined on a separable metric space. In our framework, the estimation needs only one observation of the process within a long time. Our approach is based on a generalization of the multiplicative intensity model, introduced by Aalen in the seventies. We provide consistent estimators of these two functions, under some assumptions related to the ergodicity of an embedded chain and the characteristics of the process. A numerical example illustrates our theoretical results.

It has been published in Ann. Inst. H. Poincaré Probab. Statist. [16].

6.2. Nonparametric estimation of the conditional distribution of the inter-jumping times for piecewise-deterministic markov processes

Participants: Romain Azais, François Dufour, Anne Gégout-Petit.

Piecewise-deterministic Markov process, ergodicity of Markov chains, nonparametric estimation, jump rate estimation, Nelson-Aalen estimator, asymptotic consistency

In this work, we present a nonparametric method for estimating the conditional density associated to the jump rate of a piecewise-deterministic Markov process. In our framework, the estimation needs only one observation of the process within a long time interval. Our method relies on a generalization of Aalen?s multiplicative intensity model. We prove the uniform consistency of our estimator, under some reasonable assumptions related to the primitive characteristics of the process. A simulation study illustrates the behavior of our estimator.

It has been accepted for publication in Scandinavian Journal of Statistics [17].

6.3. Finite Linear Programming Approximations of constrained discounted Markov decision processes

Participant: François Dufour.

Constrained Markov decision processes, linear programming approach to control problems, quantization, approximation of Markov decision processes

We consider a Markov decision process (MDP) with constraints under the total expected discounted cost optimality criterion. We are interested in proposing approximation methods of the optimal value of this constrained MDP. To this end, starting from the linear programming (LP) formulation of the constrained MDP (on an infinite-dimensional space of measures), we propose a finite state approximation of this LP problem. This is achieved by suitably approximating a probability measure underlying the random transitions of the dynamics of the system. Explicit convergence orders of the approximations of the optimal constrained cost are obtained. By exploiting convexity properties of the class of relaxed controls, we reduce the LP formulation of the constrained MDP to a finite-dimensional static optimization problem, that can be used to obtain explicit numerical approximations of the corresponding optimal constrained cost. A numerical application illustrates our theoretical results.

These results have been obtained in collaboration with Tomas Prieto-Rumeau, Department of Statistics and Operations Research, UNED, Madrid, Spain.

It has been published in SIAM Journal of Control and Optimization [25].

6.4. Stochastic Approximations of Constrained Discounted Markov Decision Processes

Participant: François Dufour.

Constrained Markov decision processes; Linear programming approach to control problems; Approximation of Markov decision processes

We consider a discrete-time constrained Markov decision process under the discounted cost optimality criterion. The state and action spaces are assumed to be Borel spaces, while the cost and constraint functions might be unbounded. We are interested in approximating numerically the optimal discounted constrained cost. To this end, we suppose that the transition kernel of the Markov decision process is absolutely continuous with respect to some probability measure μ. Then, by solving the linear programming formulation of a constrained control problem related to the empirical probability measure μ_n of μ, we obtain the corresponding approximation of the optimal constrained cost. We derive a concentration inequality which gives bounds on the probability that the estimation error is larger than some given constant. This bound is shown to decrease exponentially in n. Our theoretical results are illustrated with a numerical application based on a stochastic version of the Beverton-Holt population model.

These results have been obtained in collaboration with Tomas Prieto-Rumeau, Department of Statistics and Operations Research, UNED, Madrid, Spain.

It has been accepted for publication in Journal of Mathematical Analysis and Applications [26].

6.5. The expected total cost criterion for Markov decision processes under constraints

Participant: François Dufour.

Markov decision process, expected total cost criterion, constraints, linear programming, occupation measure

In this work, we study discrete-time Markov decision processes (MDPs) with constraints when all the objectives have the same form of expected total cost over the infinite time horizon. Our objective is to analyze this problem by using the linear programming approach. Under some technical hypotheses, it is shown that if there exists an optimal solution for the associated linear program then there exists a randomized stationary policy which is optimal for the MDP, and that the optimal value of the linear program coincides with the optimal value of the constrained control problem. A second important result states that the set of randomized stationary policies provides a sufficient set for solving this MDP. It is important to notice that, in contrast with the classical results of the literature, we do not assume the MDP to be transient or absorbing. More importantly, we do not impose the cost functions to be non-negative or to be bounded below. Several examples are presented to illustrate our results.

These results have been obtained in collaboration with Alexey Piunovskiy from Department. of Mathematical Sciences.

It has been published in Advances in Applied Probability [24].

6.6. Optimal stopping for piecewise-deterministic Markov processes and applications

Participants: Adrien Brandejsky, Benoîte de Saporta, François Dufour, Huilong Zhang.

We worked further on numerical methods for optimal stopping of PDMPs. On the one hand, we applied our numerical method to compute an optimal maintenance date to the test case of the heated hold-up tank. The system consists of a tank containing a fluid whose level is controlled by three components: two inlet pumps and one outlet valve. A thermal power source heats up the fluid. The failure rates of the components depends on the temperature, the position of the three components monitors the liquid level in the tank and the liquid level determines the temperature. Therefore, this system can be modeled by a hybrid process where the discrete (components) and continuous (level, temperature) parts interact in a closed loop. We model the system by a piecewise deterministic Markov process, propose and implement a numerical method to compute the optimal maintenance date to repair the components before the total failure of the system. This work is published in [30].

On the other hand, we investigated the optimal stopping problem under partial observations for PDMPs. We first obtain a recursive formulation of the optimal filter process and derive the dynamic programming equation of the partially observed optimal stopping problem. Then, we propose a numerical method, based on the quantization of the discrete-time filter process and the inter-jump times, to approximate the value function and to compute an ϵ-optimal stopping time. We prove the convergence of the algorithms and bound the rates of convergence. This work is published in [20].

6.7. Stochastic control for underwater optimal trajectories

Participants: Benoîte de Saporta, François Dufour, Huilong Zhang.

This work aims to compute optimal trajectories for underwater vehicles evolving in a given environment to accomplish some tasks. This is an optimal control problem. In real context, available inputs are not perfectly known. Hence a stochastic approach seems to be needed, coupled with the outputs of the tracking algorithms. Markov decision processes (MDPs) constitute a general family of controlled stochastic processes suitable for the modeling of sequential decision-making problems. The analysis of MDPs leads to mathematical and computational problems. The corresponding theory has reached a rather high degree of maturity, although the classical tools (such as value iteration, policy iteration, linear programming, and their various extensions) are generally hardly applicable in practice. Hence, solving MDPs numerically is an awkward and important problem. The method is applied to control a submarine which wants to well detect one or several targets and only has the information given by the tracking algorithms from the sonar observations [47].

6.8. Modeling of cell division data

Participants: Benoîte de Saporta, Anne Gégout-Petit.

This work is in collaboration with Laurence Marsalle (Univ. Lille 1).

A rigorous methodology is proposed to study cell division data consisting in several observed genealogical trees of possibly different shapes. The procedure takes into account missing observations, data from different trees, as well as the dependence structure within genealogical trees. Its main new feature is the joint use of all available information from several data sets instead of single data set estimation, to avoid the drawbacks of low accuracy for estimators or low power for tests on small single trees. The data is modeled by an asymmetric bifurcating autoregressive process and possibly missing observations are taken into account by modeling the genealogies with a two-type Galton-Watson process. Least-squares estimators of the unknown parameters of the processes are given and symmetry tests are derived. Results are applied on real data of Escherichia coli division and an empirical study of the convergence rates of the estimators and power of the tests is conducted on simulated data. This work is to appear in [29].

We have also presented a new model of asymmetric bifurcating autoregressive process with random coefficients. We couple this model with a Galton-Watson tree to take into account possibly missing observations. We propose least-squares estimators for the various parameters of the model and prove their consistency, with a convergence rate, and asymptotic normality. We use both the bifurcating Markov chain and martingale approaches and derive new results in both these frameworks. This work is to appear in [28].

6.9. Numerical method for the filtering of Markov jump linear systems

Participants: Benoîte de Saporta, Eduardo Costa.

We are interested in efficient pre-computations of the solutions of Markov switching Riccati equations. These equations are matrix-valued and naturally arise in control or filtering problems for Markov jump linear systems. It is crucial for applications to be able to compute the filter in real time, although the solutions to Riccati equations are slow to compute. Hence the need for pre-computations, taking into account the random possible changes of regimes. We propose a numerical method based on the discretization by quantization of the underlying Markov chain.

6.10. Optimization of the assembly line of the future European launcher

Participants: Benoîte de Saporta, François Dufour, Christophe Nivot.

In collaboration with Astrium space transportation, we have started working on the optimization of the assembly line of the future European launcher. We have started with a simplified model with five components to be assembled in workshops liable to breakdowns. We have modeled the problem using the Markov Decision Processes (MDP) framework and built a simulator of the process in order to run an optimization procedure

6.11. A variable clustering approach for the typology of units: a survey on farming and environment

Participants: Jérôme Saracco, Marie Chavent.

A survey on farming and environment dealing with the current transformations of the farmer job is considered. We propose to replace the usual data mining strategy which consists of applying Multiple Correspondence Analysis by a variable clustering approach. Clustering of variables aims at lumping together variables which are strongly related to each other and thus bring the same information. The ClustOfVar approach used in this paper provides at the same time groups of variables and their associated synthetic variables. In this algorithm, the homogeneity criterion of a cluster is defined by the squared Pearson correlation for the quantitative variables and by the correlation ratio for the qualitative variables. The step of variable clustering enables to get synthetic variables that can be read as a gradient. In our case study, values correspond to some relevant groupings of categories. This enables to interpret and name easily the synthetic variables. Trends in the opinion of farmers are thus highlighted with the variable clustering approach. Then we clarify these first results by applying a clustering method on the scores of the individuals measured by the synthetic variables. At the sociological level, the supply provided by the synthetic variables to interpret the clusters of farmers is obvious.

These results have been obtained in collaboration with Vanessa Kuentz from Irstea, UR ADBX.

They have been published in Journal de la Société Française de Statistique [31].

6.12. Multiple Facctorial Analysis for mixed data type

Participants: Jérôme Saracco, Marie Chavent, Amaury Labenne.

Multiple Factor Analysis (MFA) originally proposed by Escofier and Pages in 1982 is a method dedicated to the study of a set of n individuals described by groups of quantitative variables. Later, this method was extended to take into account groups of qualitative variables (Pages, 1983) then simultaneously quantitative groups and qualitative groups (Pages, 2002). However, this method does not currently take into account mixed groups, that is to say containing both quantitative and qualitative variables. The aim of our study is to propose sustainable development indicators by integrating the aspect of quality of life. For that, we are confronted with the analysis of groups of variables with quantitative and qualitative variables. In this work, we propose an extension of the MFA method, called MFAMIX, for the multiple factor analysis of mixed groups of variables. This approach relies on a combination of AFM and PCAMIX method that allows the analysis of mixed data. MFAMIX method is presented using a singular value decomposition and illustrated on socio-economic data about the quality of life.

These results have been obtained in collaboration with Vanessa Kuentz from Irstea, UR ADBX.

They have been have been presented in two national conferences [43], [41].

6.13. Detecting mental states of alertness with genetic algorithm variable selection

Participants: Marie Chavent, Laurent Vézard.

The objective of the present work is to develop a method able to automatically determine mental states of vigilance; i.e., a person's state of alertness. Such a task is relevant to diverse domains, where a person is expected or required to be in a particular state. For instance, pilots or medical staffs are expected to be in a highly alert state, and this method could help to detect possible problems. In this paper, an approach is developed to predict the state of alertness ("normal" or "relaxed") from the study of electroencephalographic signals (EEG) collected with a limited number of electrodes. The EEG of 58 participants in the two alertness states (116 records) were collected via a cap with 58 electrodes. After a data validation step, 19 subjects were retained for further analysis. A genetic algorithm was used to select an optimal subset of electrodes. Common spatial pattern (CSP) coupled to linear discriminant analysis (LDA) was used to build a decision rule and thus predict the alertness of the participants. Different subset sizes were investigated and the best result was obtained by considering 9 electrodes (correct classification rate of 73.68

These results have been obtained in collaboration with Pierrick Legrand from Alea Inria team and Leonardo Trujillo from Instituto Tecnologico de Tijuana.

This work has been presented in a international IEEEl conference [38].

6.14. ClustOfVar : an R package for dimension reduction via clustering of variables. Application in supervised classification and variable selection in gene expressions data

Participants: Marie Chavent, Jérôme Saracco.

The main goal of this work is to tackle the problem of dimension reduction for high-dimensional supervised classication. The motivation is to handle gene expression data. The proposed method works in 2 steps. First, one eliminates redundancy using clustering of variables, based on the R-package ClustOfVar. This first step is only based on the exploratory variables (genes). Second, the synthetic variables (summarizing the clusters obtained at the first step) are used to construct a classifier (e.g. logistic regression, LDA, random forests). We stress that the first step reduces the dimension and gives linear combinations of original variables (synthetic variables). This step can be considered as an alternative to PCA. A selection of predictors (synthetic variables) in the second step gives a set of relevant original variables (genes). Numerical performances of the proposed procedure are evaluated on gene expression datasets. We compare our methodology with LASSO and sparse PLS discriminant analysis on these datasets.

This work is a collaboration with Robin Genuer from SISTM Inria team and Benoit Liquet from University of Queensland.

This work has been presented in a international workshop on Statistical Methods for (post)-Genomics Data (SMPGD 2013) [42].

6.15. A sliced inverse regression approach for data stream

Participants: Jérôme Saracco, Marie Chavent.

This work is in collaboration with Stéphane Girard (Inria Grenoble Alpes), Benoît Liquet (MRC, Cambridge University), Vanessa Kuentz (Irstea) and Thi Mong Gnoc Nguyen (Univ. de Strasbourg).

In this work, we focus on data arriving sequentially by blocks in a stream. A semiparametric regression model involving a common EDR (Effective Dimension Reduction) direction is assumed in each block. Our goal is to estimate this direction at each arrival of a new block. A simple direct approach consists of pooling all the observed blocks and estimating the EDR direction by the SIR (Sliced Inverse Regression) method. But in practice, some disadvantages appear such as the storage of the blocks and the running time for large sample sizes. To overcome these drawbacks, we propose an adaptive SIR estimator of based on the optimization of a quality measure. The corresponding approach is faster both in terms of computational complexity and running time, and provides data storage benefits. The consistency of our estimator is established and its asymptotic distribution is given. An extension to multiple indices model is proposed. A graphical tool is also provided in order to detect changes in the underlying model, i.e., drift in the EDR direction or aberrant blocks in the data stream. A simulation study illustrates the numerical behavior of our estimator. Finally, an application to real data concerning the estimation of physical properties of the Mars surface is presented.

This work is to appear in [21].

6.16. Comparison of Kernel Density Estimators with Assumption on Number of Modes

Participant: Jérôme Saracco.

This work is in collaboration with Bernard Bercu (Univ. Bretagne Sud) and Thi Mong Gnoc Nguyen (Univ. de Strasbourg).

In this work, we investigate the asymptotic behavior of the Nadaraya-Watson estimator for the estimation of the regression function in a semiparametric regression model. On the one hand, we make use of the recursive version of the sliced inverse regression method for the estimation of the unknown parameter of the model. On the other hand, we implement a recursive Nadaraya-Watson procedure for the estimation of the regression function which takes into account the previous estimation of the parameter of the semiparametric regression model. We establish the almost sure convergence as well as the asymptotic normality for our Nadaraya-Watson estimator. We also illustrate our semiparametric estimation procedure on simulated data.

This work is to appear in [19].

6.17. Comparison of Kernel Density Estimators with Assumption on Number of Modes

Participants: Jérôme Saracco, Raphaël Coudret.

This work is in collaboration with Gilles Durrieu (Univ. Bretagne Sud).

A data-driven bandwidth choice for a kernel density estimator called critical bandwidth is investigated. This procedure allows the estimation to have as many modes as assumed for the density to estimate. Both Gaussian and uniform kernels are considered. For the Gaussian kernel, asymptotic results are given. For the uniform kernel, an argument against these properties is mentioned. These theoretical results are illustrated with a simulation study which compare the kernel estimators that rely on critical bandwidth with another one which uses a plug-in method to select its bandwidth. An estimator that consists in estimates of density contour clusters and takes assumptions on number of modes into account is also considered. Finally, the methodology is illustrated using environment monitoring data.

This work is to appear in [22].

6.18. Comparison of sliced inverse regression approaches for underdetermined cases

Participants: Jérôme Saracco, Raphaël Coudret.

This work is in collaboration with Benoît Liquet (MRC, Cambridge University).

Among methods to analyze high-dimensional data, the sliced inverse regression (SIR) is of particular interest for non-linear relations between the dependent variable and some indices of the covariate. When the dimension of the covariate is greater than the number of observations, classical versions of SIR cannot be applied. Various upgrades were then proposed to tackle this issue such as regularized SIR (RSIR) and sparse ridge SIR (SR-SIR), to estimate the parameters of the underlying model and to select variables of interest. In this paper, we introduce two new estimation methods respectively based on the QZ algorithm and on the Moore-Penrose pseudo-inverse. We also describe a new selection procedure of the most relevant components of the covariate that relies on a proximity criterion between submodels and the initial one. These approaches are compared with RSIR and SR-SIR in a simulation study. Finally we applied SIR-QZ and the associated selection procedure to a genetic dataset in order to find markers that are linked to the expression of a gene. These markers are called expression quantitative trait loci (eQTL).

This work was presented in a national conference [23] and is to appear in [37].

6.19. Conditional Quantile Estimationthrough Optimal Quantization

Participants: Jérôme Saracco, Isabelle Charlier.

This work is in collaboration with Davy Paindaveine (Univ. Libre de Bruxelles).

In this work, we construct a nonparametric estimator of conditional quantiles of Y given X via optimal quantization. In a first step, we propose to approximate conditional quantiles thanks to optimal quantization in L^p-norm, consisting in discretizing X and Y thanks to some optimal grids of size N. We state a result of convergence of this approximation toward the true conditional quantile. The estimator was implemented in **R** in order to evaluate its numerical behavior and to compare it with existing estimators. A simulation study illustrates the good behavior of our estimator. The practical choice of N is discussed. We apply our approach to a real data set.

This work was presented in a national conference [35].

6.20. Estimation of water consumption based on survey techniques using an automatic meter reading sample

Participant: Jérôme Saracco.

This work is in collaboration with Karim Claudio (LyRE), Vincent Couallier (Univ. de Bordeaux) and Yves Le Gat (Irstea).

Automatic water meters reading are, nowadays, the best technology for real time knowledge of water consumption. At an hydraulic sector scale, a complete equipment permits to know the total consumption of a finite size population, for a time scale as small as the hour. However its cost for generalization is sometimes unbearable for the collectivity, for whom sampling techniques have to be set up. In a objective of a total consumption estimation, this article describes and compares standard methods of survey techniques and propose to retain a methodology for implementation of an operational sample and to calibrate the corresponding total estimator.

This work was presented in a national conference [36] and an associated paper is currently in revision.

6.21. Hidden Markov Model for the detection of a degraded operating mode of optronic equipment

Participants: Camille Baysse, Anne Gégout-Petit, Jérôme Saracco.

This work is in collaboration with Didier Bihannic (Thales Optronics) and Michel Prenat (Thales Optronics).

As part of optimizing the reliability, Thales Optronics now includes systems that examine the state of its equipment. The aim of this work is to use hidden Markov Model to detect as soon as possible a change of state of optronic equipment in order to propose maintenance before failure. For this, we carefully observe the dynamic of a variable called "cool down time" and noted Tmf, which reflects the state of the cooling system. Indeed, the Tmf is an observation of the hidden state of the system. This one is modeled by a Markov chain and the Tmf is a noisy function of it. Thanks to filtering equations, we obtain results on the probability that an appliance is in degraded state at time t, knowing the history of the Tmf until this moment. We have evaluated the numerical behavior of our approach on simulated data. Then we have applied this methodology on our real data and we have checked that the results are consistent with the reality. This method can be implemented in a HUMS (Health and Usage Monitoring System). This simple example of HUMS would allow the Thales Optronics Company to improve its maintenance system. This company will be able to recall appliances which are estimated to be in degraded state and do not control too early those estimated in stable state.

This work is to appear in [18].

6.22. A new sliced inverse regression method for multivariate response

Participants: Jérôme Saracco, Raphaël Coudret.

This work is in collaboration with Stéphane Girard (Inria Grenoble Alpes).

We consider a semiparametric regression model of a q-dimensional multivariate response y on a p-dimensional covariate x. In this paper, a new approach is proposed based on sliced inverse regression for estimating the effective dimension reduction (EDR) space without requiring a prespecified parametric model. The convergence at rate $\sqrt{n}$ of the estimated EDR space is shown. We discuss the choice of the dimension of the EDR space. The numerical performance of the proposed multivariate SIR method is illustrated on a simulation study. Moreover, we provide a way to cluster components of y related to the same EDR space. One can thus apply properly multivariate SIR on each cluster instead of blindly applying multivariate SIR on all components of y. An application to hyperspectral data is provided.

This work is currently under revision, see [48].

6.23. An introduction to dimension reduction in nonparametric kernel regression

Participant: Jérôme Saracco.

This work is in collaboration with Stéphane Girard (Inria Grenoble Alpes).

Nonparametric regression is a powerful tool to estimate nonlinear relations between some predictors and a response variable. However, when the number of predictors is high, nonparametric estimators may suffer from the curse of dimensionality. In this chapter, we show how a dimension reduction method (namely Sliced Inverse Regression) can be combined with nonparametric kernel regression to overcome this drawback. The methods are illustrated both on simulated datasets as well as on an astronomy dataset using the **R** software.

This work was presented in "School in Astrostatistics" (Annecy, October, 21-25, 2013) and is to appear as a chapter in book intilted *Methods and Applications of Regression in Astrophysics* in 2014.

7. Bilateral Contracts and Grants with Industry

7.1. Astrium

Participants: Romain Azaïs, Adrien Brandejsky, Benoîte de Saporta, François Dufour, Anne Gégout-Petit, Christophe Nivot, Huilong Zhang.

The goal of this project is to propose models for fatigue of structure and to study an approach to evaluate the probability of occurrence of events defined by the crossing of a threshold. In this context, Astrium funded the PhD Thesis of Adrien Brandejsky (2009-2012) and is a partner of ANR Fautocoes. A new contract started in 2013 about the optimization of the assembly line of the future European launcher.

7.2. DCNS

Participants: Benoîte de Saporta, François Dufour, Huilong Zhang.

In september 2010, an industrial collaboration started with DCNS on the application of Markov Decision Processes to optimal stochastic control of a submarine to maximize the acoustic signature of a target vessel. In 2012, we extended our previous results to multiple target vessels and 3D control. We also coupled our code with the output of a tracking software to take more realistically into account the uncertainty on the position and speed of the targets. In 2013, we coupled our optimization procedure with the output of the tracking algorithms to estimate the positions of the targets.

7.3. Thales Optronique

Participants: Camille Baysse, Benoîte de Saporta, François Dufour, Anne Gégout-Petit, Jérôme Saracco.

Integrated maintenance, failure intensity, optimisation.

As part of optimizing the reliability, Thales Optronics includes systems that examine the state of their equipment . This function is performed by HUMS (Health Unit Monitoring Systems). The collaboration is the subject of the PhD of Camille Baysse (CIFRE). The aim of this thesis is to implement in the HUMS a program based on observations that can determine the state of the system, optimize maintenance operations and evaluate the failure risk of a mission. This work was presented in the conferences [33], [40] and is to appear in [18].

8. Partnerships and Cooperations

8.1. Regional Initiatives

8.1.1. PSI : Psychology and Sound Interactions

The aim of this project was to develop a classifier to automatically determine the alertness state of humans from electroencephalographic (EEG) signals. Such a task is relevant to diverse domains, where a person is expected to be in a highly alert state. The goal was to contrust a Brain-Computer Interface (BCI) based on synthetized music to modify alterness state of a person. This Région Aquitaine grant (2010-2013) grant included the PHD-grant of Laurent Vezard.

8.1.2. Chaire Inria-Astrium-EADS IW-Conseil régional d'Aquitaine

The chaire is funding the PhD thesis of Christophe Nivot on the optimization of the assembly line of the future European launcher.

8.2. National Initiatives

8.2.1. ANR FAUTOCOES

The goal of the project "FAUTOCOES" (number ANR-09-SEGI-004) of the ARPEGE program of the French National Agency of Research (ANR) can be described as follows. Today, complex technological processes must maintain an acceptable behavior in the event of random structural perturbations, such as failures or component degradation. Aerospace engineering provides numerous examples of such situations: an aircraft has to pursue its mission even if some gyroscopes are out of order, a space shuttle has to succeed in its re-entry trip with a failed on-board computer. Failed or degraded operating modes are parts of an embedded system history and should therefore be accounted for during the control synthesis.

These few basic examples show that complex systems like embedded systems are inherently vulnerable to failure of components and their reliability has to be improved through fault-tolerant control. Embedded systems require mathematical representations which are in essence dynamic, multi-model and stochastic. This increasing complexity poses a genuine scientific challenge:

- to model explicitly and realistically the dynamical interactions existing between the physical state variables defining the system: pressure, temperature, flow rate, intensity, etc, and the functional and dysfunctional behavior of its components;
- to estimate the performance of the system through the evaluation of reliability indexes such as availability, quality, and safety;
- to optimize the control to prevent system failures, as well as to maintain the system function when a failure has occurred.

Our aim is to meet the previously mentioned challenge by using the framework of piecewise deterministic Markov processes (PDMP's in short) with an emphasis on probabilistic and deterministic numerical methods. More precisely, our objectives are

- to use the framework of piecewise deterministic Markov processes to model complex physical systems and phenomena;
- to compute expectations of functionals of the process in order to evaluate the performance of the system;
- to develop theoretical and numerical control tools for PDMP's to optimize the performance and/or to maintain system function when a failure has occurred.

More details are available at http://fautocoes.bordeaux.inria.fr/.

8.2.2. ANR ADAPTEAU

The ANR project ADAPTEAU has been obtained for the period 2012-2016 and will start in january 2012.

ADAPTEAU aims to contribute to the analysis and management of global change impacts and adaptation patterns in River-Estuarine Environments (REEs) by interpreting the scientific challenges associated with climate change in terms of: i) scale mismatches; ii) uncertainty and cognitive biases between social actors; iii) interdisciplinary dialogue on the "adaptation" concept; iv) critical insights on adaptive governance and actions, v) understanding the diversity of professional, social and economic practices vis-à-vis global change. The project aims to build an integrative and interdisciplinary framework involving biophysical and social sciences, as well as stakeholders and civil society partners. The main objective is to identify adaptive strategies able to face the stakes of global change in REEs, on the basis of what we call 'innovative adaptation options'.

We consider the adaptation of Social-Ecological Systems (SES) through the expected variations of the hydrological regimes (floods / low-flow) of the Garonne-Gironde REE—a salient issue in SW France, yet with a high potential for genericity The ADAPTEAU project will be organised as follows:

- Achieve and confront socio-economic and environmental assessments of expected CC impacts on the Garonne-Gironde river-estuarine continuum (task 1);
- Identify the emerging 'innovative adaptation options' endorsed by various social, economic, political actors of the territory (depolderisation, 'room for rivers' strategies, changes in economic activities, agricultural systems or social practices), then test their environmental, economic and social robustness through a selected subset (task 2);
- Scientists, representatives from administrators and civil society collaborate to build adaptation scenarios, and discuss them in pluralistic arenas in order to evaluate their social and economic feasibility, as well as the most appropriate governance modes (task 3).
- Disseminate the adaptation strategies to academics and managers, as well as to the broader society (task 4).

  The expected results are the definition and diffusion of new regional-scale reference frameworks for the discussion of adaptation scenarios in REE and other SESs, as well as action guidelines to better address climate change stakes.

The CQFD team work on tasks 1 and 3.

8.2.3. ANR Piece

ANR Piece (2013-2016) of the program *Jeunes chercheuses et jeunes chercheurs* of the French National Agency of Research (ANR), lead by F. Malrieu (Univ. Tours). The 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 of complex 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, modeling) 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.

8.3. European Initiatives

8.3.1. Collaborations in European Programs, except FP7

Numerical methods for Markov decision processes (2013-2015) This project is funded by the Gobierno de Espana, Derccion Genral de Investigacion Cinetifica y Tecnica (reference number: MTM2012-31393) for three years to support the scientific collaboration between Tomas Prieto-Rumeau and François Dufour. This research project is concerned with numerical methods for Markov decision processes (MDPs). Namely, we are interested in approximating numerically the optimal value function and the optimal controls for different classes of constrained and unconstrained MDPs. Our methods are based on combining the linear programming formulation of an MDP with a discretization procedure —referred to as quantization— of a probability distribution, underlying the random transitions of the dynamic system. We are concerned with optimality criteria such as the total expected cost criterion (for finite horizon problems) and, on the other hand, the total expected discounted cost and the average cost optimality criteria (for infinite horizon problems).

8.4. International Initiatives

8.4.1. Participation In other International Programs

Control of Dynamic Systems Subject to Stochastic Jumps USP-COFECUB grant (2013-2016). The main goals of this joint cooperation will be to study the control of dynamic systems subject to stochastic jumps. Three topics will be considered throughout the next 3 years. In the first topic we will study the control problem of piecewise-deterministic Markov processes (PDMP's) considering constraints. In this case the main goal is to obtain a theoretical formulation for the equivalence between the original optimal control of PDMP's with constrains and an infinite dimensional static linear optimization problem over a space of occupation measures of the controlled process. F. Dufour at Inria and O. Costa in USP will mainly carry out this topic. In the second topic we will focus on numerical methods for solving control and filtering problems related to Markov jump linear systems (MJLS). This project will allow a first cooperation between B. de Saporta and E. Costa. The third research subject will be focused on quantum control by using Lyapunov-like stochastic methods and P. Rouchon and P. Pereira da Silva will conduct it.

8.5. International Research Visitors

8.5.1. Visits of International Scientists

Eduardo Costa (Univ. São Paulo), invited from July 22nd to August 1st 2013, USP-COFECUB grant.

8.5.2. Visits to International Teams

Benoîte de Saporta was invited one week (April 22-April 29) by Jian-Fang Yao at the University of Hong Kong. Benoîte de Saporta was invited three weeks (May 22-June 8) by Eduardo Costa at the University of São Paulo in São Carlos, Brazil (USP-COFECUB grant).

9. Dissemination

9.1. Scientific Animation

The team CQFD organized a *journée SMAI maths-industrie* on dependability and safety in Bordeaux in April 2013.

The CQFD team organized the workshop Statlearn'13 on the challenging problems in Statistical Learning in Bordeaux on April 8-9, 2013.

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

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

All the member of the team are regular reviewers for the most important journals in applied probability and statistics.

F. Dufour is member of the scientific council of the engineering school ENSEIRB-MATMECA.

F. Dufour is member of the scientific council of the Institute of Mathematics of Bordeaux.

F. Dufour was vice-president of the Inria Project Committee.

B. de Saporta is president of the "Congress and Colloquium" commission of the Inria Bordeaux Sud-Ouest.

B. de Saporta is in charge of the seminar of the team *Statistics and Probability* of the Institute of Mathematics of Bordeaux (IMB).

B. de Saporta is a member of the council of the Institute of Mathematics of Bordeaux (IMB).

A. Gégout-Petit was elected member of the CEVU of University Bordeaux Segalen

A. Gégout-Petit was member of the Mathematical Institute of Bordeaux council

A. Gégout-Petit was general secretary and elected member of the council of the Société Française de Statistique.

J. Saracco is member of the commission Inria ”Jeunes Chercheurs”.

J. Saracco is member of the council of ENSC

J. Saracco is the leader of the team ”Statistics and Probability” of the Institute of Mathematics of Bordeaux (IMB).

B de Saporta, J. Saracco and M. Chavent are elected members of the CNU 26.

J. Saracco is the head of the engineering department of ENSC.

F. Dufour is the head of the second year cursus engineering school MATMECA.

M. Chavent is co-director of the cursus *Modélisation Statistique et Sochastique* of the master MIMSE *Ingénierie Mathématique, Statistique et Economique* of the University of Bordeaux.

H. Zhang is director of the cursus *Ingénierie Mathématique* of the Licence de Mathématiques of the University of Bordeaux.

9.2. Teaching - Supervision - Juries

9.2.1. Teaching

Licence : F. Dufour, Probabilités et statistiques, 16 heures, niveau L3, Institut Polytechnique de Bordeaux, école ENSEIRB-MATMECA, France.

Licence : F. Dufour, Probabilités 10,6 heures, niveau L3, Institut Polytechnique de Bordeaux, école ENSEIRB-MATMECA, France.

Master : F. Dufour, Méthodes numériques pour la fiabilité, 24 heures, niveau M1, Institut Polytechnique de Bordeaux, école ENSEIRB-MATMECA, France.

Master : F. Dufour, Probabilités, 20 heures, niveau M1, Institut Polytechnique de Bordeaux, école ENSEIRB-MATMECA, France.

Licence : A. Gégout-Petit, Etudes de cas en statistique, 28h, L3 MASS (applied mathematics), Université Bordeaux Segalen, France.

Licence : A. Gégout-Petit, Econométrie et séries chronologiques, 24h, L3 MASS (applied mathematics), Université Bordeaux Segalen, France.

Master : A. Gégout-Petit, Analyse de variance, 36h, M1, université Bordeaux, France.

Licence : M. Chavent, Statistique descriptive, 36 ETD ,L1, university Bordeaux Segalen, France

Master : M. Chavent, Analyse des données 2, 25 ETD, niveau M2, university Bordeaux Segalen, France

Master : M. Chavent, Scoring, 21 ETD, niveau M2, university Montesquieu Bordeaux 4, France

Licence: J. Saracco, Descriptive statistics, 10.5h, L3, First year of ENSC, France

Licence: J. Saracco, Mathematical statistics, 20h, L3, First year of ENSC, France

Licence:J. Saracco, Data analysis (multidimensional statistics), 20h, L3, First year of ENSC, France

Licence: J. Saracco, Mathematics (complement of linear algebra), 20h, L3, First year of ENSC, France

Master: J. Saracco, Statistical modeling, 20h, M1, Second year of ENSC, France

Master: J. Saracco, training project, 20h, M1, Second year of ENSC, France

Master : B. de Saporta, Processus aléatoires en finance 30h ETD, M1, université de Bordeaux, France

Master : B. de Saporta, Finance en temps continu, 10h ETD, M2, université de Bordeaux, France

Master : B. de Saporta, Finance en temps discret, 29h ETD, M2, université de Bordeaux, France

Master : B. de Saporta, Processus de Markov, 25h ETD, M2, université de Bordeaux, France

9.2.2. Supervision

HdR : B. de Saporta, Contribution à l'estimation et au contrôle de processus stochastiques, Univ. Bordeaux I, July 2013

PhD in progress : Christophe Nivot, Optimisation de la chaîne de montage du futur lanceur européen, sept. 2013, B. de Saporta and F. Dufour

PhD completed : Azaïs Romain, Inférence des processus Markoviens déterministe par morceaux , juillet 2013, supervised by François Dufour and Anne Gégout-Petit

PhD completed : Camille Baysse, Analyse et optimisation de la fiabilité d'un équipement opto-électronique équipé de HUMS, novembre 2013, supervised by Anne Gégout-Petit and Jérôme Saracco

PhD completed : Laurent Vezard, Classification de signaux EEG et synthèse de paramètres musicaux par algorithme évolutionnaire, december 2013, supervised by M. Chavent and P. Legrand.

PhD completed : Raphaël Coudret, Modélisation statistique de données acquises à haute fréquence : application en environnement et génétique, september 2013, supervised by J. Saracco and G.Durrieu.

PhD in progress : Karim Claudio, Un outil d'aide à la maîtrise des pertes dans les réseaux d'eau potable : mise en place d'un modèle de fuite multi-état en secteur hydraulique instrumenté , supervised by J. Saracco and V. Couallier.

PhD in progress : Amaury Labenne, Approche Statistique du diagnostic territorial par la notion de qualité de vie, supervised by M. Chavent, J. Saracco and V. Kuentz.

PhD in progress : Adrien Todeschini, Elaboration et validation d'un système de recommandation bayésien, supervised by F. Caron and M. Chavent.

PhD in progress : Isabelle Charlier, Optimal quantization applied to conditional quantile estimation, University of Bordeaux 1 and Université Libre de Bruxelle, supervised by J. Saracco and D. Paindaveine.

9.2.3. Juries

B. de Saporta was a member of the PhD jury of Camille Baysse. B. de Saporta is an elected member of CNU 26.

9.3. Popularization

B. de Saporta took part in a speed mediation event at Inria Bordeaux Sud Ouest (dec. 2013).

10. Bibliography

Major publications by the team in recent years

[1] M. CHAVENT, B. LIQUET, J. SARACCO. *A semiparametric approach for a multivariate sample selection model*, in "Statist. Sinica", 2010, vol. 20, n^{o} 2, pp. 513–536

[2] O. COSTA, F. DUFOUR. *Stability and ergodicity of piecewise deterministic Markov processes*, in "SIAM J. Control Optim.", 2008, vol. 47, n^{o} 2, pp. 1053–1077

[3] O. COSTA, F. DUFOUR. *The Vanishing Discount Approach for the Average Continuous Control of Piecewise Deterministic Markov Processes*, in "Journal of Applied Probability", 2009, vol. 46, n^{o} 4, pp. 1157-1183

[4] F. DUFOUR, A. PIUNOVSKIY. *Multi-objective stopping problem for discrete-time Markov processes*, in "Journal of Applied Probability", 2010, vol. 47, n^{o} 4, pp. 947-966

[5] A. GANNOUN, J. SARACCO, A. YUAN, G. E. BONNEY. *Non-parametric quantile regression with censored data*, in "Scand. J. Statist.", 2005, vol. 32, n^{o} 4, pp. 527–550

[6] J. SARACCO. *Asymptotics for pooled marginal slicing estimator based on* SIR_{α} *approach*, in "J. Multivariate Anal.", 2005, vol. 96, n^{o} 1, pp. 117–135

[7] B. DE SAPORTA, F. DUFOUR, H. ZHANG, C. ELEGBEDE. *Optimal stopping for predictive maintenance of a structure subject to corrosion*, in "Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability", 2012, vol. 226, n^{o} 2, pp. 169-181, http://hal.inria.fr/hal-00554759

[8] B. DE SAPORTA, A. GÉGOUT-PETIT, L. MARSALLE. *Parameters estimation for asymmetric bifurcating autoregressive processes with missing data*, in "Electronic Journal of Statistics", 2011, vol. 5, pp. 1313-1353 [*DOI :* 10.1214/11-EJS643], http://hal.inria.fr/hal-00545447/en

[9] B. DE SAPORTA, F. DUFOUR, K. GONZALEZ. *Numerical method for optimal stopping of piecewise deterministic Markov processes*, in "The Annals of Applied Probability", 2010, vol. 20, n^{o} 5, pp. 1607-1637

Publications of the year

Doctoral Dissertations and Habilitation Theses

[10] R. AZAÏS. , *Estimation non paramétrique pour les processus markoviens déterministes par morceaux*, Université Sciences et Technologies - Bordeaux I, July 2013, http://hal.inria.fr/tel-00844395

[11] C. BAYSSE. , *Analyse et optimisation de la fiabilité d'un équipement opto-électronique équipé de HUMS*, Université Sciences et Technologies - Bordeaux I, November 2013, http://hal.inria.fr/tel-00938744

[12] R. COUDRET. , *Stochastic modelling using large data sets : applications in ecology and genetics*, Université Sciences et Technologies - Bordeaux I, September 2013, http://hal.inria.fr/tel-00865867

[13] L. VEZARD. , *Réduction de dimension en apprentissage supervisé. Application à l'étude de l'activité cérébrale*, Université Sciences et Technologies - Bordeaux I, December 2013, http://hal.inria.fr/tel-00926845

[14] B. DE SAPORTA. , *Contribution à l'estimation et au contrôle de processus stochastiques*, Université Sciences et Technologies - Bordeaux I, July 2013, Habilitation à Diriger des Recherches, http://hal.inria.fr/tel-00905873

Articles in International Peer-Reviewed Journals

[15] R. AZAÏS. *A recursive nonparametric estimator for the transition kernel of a piecewise-deterministic Markov process*, in "ESAIM: Probability and Statistics", 2014 [*DOI :* 10.1051/PS/2013054], http://hal.inria.fr/hal-00759065

[16] R. AZAÏS, F. DUFOUR, A. GÉGOUT-PETIT. *Nonparametric estimation of the jump rate for non-homogeneous marked renewal processes*, in "Annales de l'Institut Henri Poincaré (B) Probabilités et Statistiques", 2013, vol. 49, n^{o} 4, pp. 1204-1231, http://hal.inria.fr/hal-00759062

[17] R. AZAÏS, F. DUFOUR, A. GÉGOUT-PETIT. *Nonparametric estimation of the jump rate for piecewise-deterministic Markov processes*, in "Scandinavian Journal of Statistics", 2014, 26 p. , To appear, http://hal.inria.fr/hal-00925874

[18] C. BAYSSE, D. BIHANNIC, A. GÉGOUT-PETIT, M. PRENAT, J. SARACCO. *Hidden Markov Model for the detection of a degraded operating mode of optronic equipment*, in "Journal de la Société Française de Statistique", 2014, 11 p. , A paraître, http://hal.inria.fr/hal-00763963

[19] B. BERCU, T. MONG NGOC NGUYEN, J. SARACCO. *On the asymptotic behavior of the Nadaraya-Watson estimator associated with the recursive SIR method*, in "Statistics", 2014, http://hal.inria.fr/hal-00673832

[20] A. BRANDEJSKY, B. DE SAPORTA, F. DUFOUR. *Optimal stopping for partially observed piecewise-deterministic Markov processes*, in "Stochastic Processes and their Applications", 2013, vol. 123, pp. 3201-3238 [*DOI :* 10.1016/J.SPA.2013.03.006], http://hal.inria.fr/hal-00755052

[21] M. CHAVENT, S. GIRARD, V. KUENTZ-SIMONET, B. LIQUET, T. M. N. NGUYEN, J. SARACCO. *A sliced inverse regression approach for data stream*, in "Computational Statistics", 2014, vol. to appear, http://hal.inria.fr/hal-00688609

[22] R. COUDRET, G. DURRIEU, J. SARACCO. *Comparison of kernel density estimators with assumption on number of modes*, in "Communication in Statistics - Simulation and Computation", October 2013, in press [*DOI :* 10.1080/03610918.2013.770530], http://hal.inria.fr/hal-00748790

[23] R. COUDRET, B. LIQUET, J. SARACCO. *Comparison of sliced inverse regression approaches for underdetermined cases*, in "Journal de la Société Française de Statistique", 2013, In press, http://hal.inria.fr/hal-00768352

[24] F. DUFOUR, A. PIUNOVSKIY. *The expected total cost criterion for Markov decision processes under constraints*, in "Advances in Applied Probability", 2013, vol. 45, n^{o} 3, pp. 837-859, http://hal.inria.fr/hal-00925859

[25] F. DUFOUR, T. PRIETO-RUMEAU. *Finite Linear Programming Approximations of constrained discounted Markov decision processes*, in "SIAM Journal of control and optimization", 2013, vol. 51, n^{o} 2, pp. 1298-1324 [*DOI :* 10.1137/120867925], http://hal.inria.fr/hal-00925862

[26] F. DUFOUR, T. PRIETO-RUMEAU. *Stochastic Approximations of Constrained Discounted Markov Decision Processes*, in "Journal of Mathematical Analysis and applications", 2014, 30 p. , To appear, http://hal.inria.fr/hal-00925878

[27] A. GÉGOUT-PETIT, L. GUÉRIN-DUBRANA, A. LABENNE, J.-C. LABROUSSE, S. BASTIEN, P. REY. *Statistical analysis of the grapevines mortality associated Esca or Eutypa dieback foliar expression*, in "Phytopathologia Mediterranea", 2013, vol. 52, n^{o} 2, pp. 276 - 288, http://hal.inria.fr/hal-00933308

[28] B. DE SAPORTA, A. GÉGOUT-PETIT, L. MARSALLE. *Random coefficients bifurcating autoregressive processes*, in "ESAIM: Probability and Statistics", 2013 [*DOI :* 10.1051/PS/2013042], http://hal.inria.fr/hal-00702357

[29] B. DE SAPORTA, A. GÉGOUT-PETIT, L. MARSALLE. *Statistical study of asymmetry in cell lineage data*, in "Computational Statistics and Data Analysis", 2014, vol. 69, pp. 15-39 [*DOI :* 10.1016/J.CSDA.2013.07.025], http://hal.inria.fr/hal-00702359

[30] B. DE SAPORTA, H. ZHANG. *Predictive maintenance for the heated hold-up tank*, in "Reliability Engineering and System Safety", 2013, vol. 115, pp. 82-90 [*DOI :* 10.1016/J.RESS.2013.02.016], http://hal.inria.fr/hal-00755056

Articles in National Peer-Reviewed Journals

[31] V. KUENTZ-SIMONET, S. LYSER, J. CANDAU, P. DEUFFIC, M. CHAVENT, J. SARACCO. *Une approche par classification de variables pour la typologie d'observations : le cas d'une enquête agriculture et environnement*, in "Journal de la Société Française de Statistique", 2013, vol. 154, n^{o} 2, 37 p. , http://hal.inria.fr/hal-00876254

Invited Conferences

[32] B. DE SAPORTA, A. BRANDEJSKY, F. DUFOUR. *Numerical method to compute the law of exit times for piecewise deterministic Markov processes*, in "Hitting times and exit problems for stochastic models", Dijon, France, 2013, http://hal.inria.fr/hal-00905866

International Conferences with Proceedings

[33] C. BAYSSE, D. BIHANNIC, A. GÉGOUT-PETIT, M. PRENAT, B. DE SAPORTA, J. SARACCO. *Maintenance Optimisation of Optronic Equipment*, in "Prognostics and System Health Management Conference", Milan, Italy, 2013, pp. 709-714, http://hal.inria.fr/hal-00906346

[34] P. MOUAWAD, M. DESAINTE-CATHERINE, A. GÉGOUT-PETIT, C. SEMAL. *The role of the singing acoustic cues in the perception of broad affect dimensions*, in "The 10th International Symposium on Computer Music Multidisciplinary Research (CMMR 2013)", Marseille, France, LABORATOIRE DE MÉCANIQUE AND D'ACOUSTIQUE (LMA) (editor), October 2013, 98 p. , http://hal.inria.fr/hal-00936752

[35] J. SARACCO, I. CHARLIER, D. PAINDAVEINE. *Application de la quantification optimale à l'estimation de quantiles conditionnels*, in "45èmes journées de statistique - Toulouse", France, 2013, http://hal.inria.fr/hal-00926466

[36] J. SARACCO, K. CLAUDIO, V. COUALLIER, Y. LE GAT. *Redressement d'un estimateur de la consommation en eau potable d'une population*, in "45èmes journées de statistique - Toulouse", France, 2013, http://hal.inria.fr/hal-00926475

[37] J. SARACCO, R. COUDRET, B. LIQUET. *Comparaison de méthodes basées sur SIR pour des cas sous-déterminés ($n < p$)*, in "45èmes journées de statistique - Toulouse", France, 2013, http://hal.inria.fr/hal-00926478

[38] L. VEZARD, P. LEGRAND, M. CHAVENT, F. FAITA-AINSEBA, L. TRUJILLO. *Detecting mental states of alertness with genetic algorithm variable selection*, in "IEEE Congress on Evolutionary Computation (CEC) 2013", CANCUN, Mexico, June 2013, pp. 1247 - 1254, http://hal.inria.fr/hal-00939851

[39] H. ZHANG, B. DE SAPORTA, F. DUFOUR, G. DELEUZE. *Dynamic Reliability by Using Simulink and Stateflow*, in "Prognostics and System Health Management Conference", Milan, Italy, 2013, vol. 33, pp. 529-534, http://hal.inria.fr/hal-00906337

National Conferences with Proceedings

[40] C. BAYSSE, D. BIHANNIC, B. DE SAPORTA, A. GÉGOUT-PETIT, M. PRENAT, J. SARACCO. *Optimisation de la maintenance d'un équipement optronique*, in "45èmes Journées de Statistique", Toulouse, France, 2013, http://hal.inria.fr/hal-00906353

[41] A. LABENNE, M. CHAVENT, V. KUENTZ-SIMONET, T. RAMBONILAZA, J. SARACCO. *Une extension de l'Analyse Factorielle Multiple pour des groupes de variables mixtes*, in "45èmes Journées de Statistique", France, 2013, http://hal.inria.fr/hal-00926254

Conferences without Proceedings

[42] M. CHAVENT, R. GENUER, V. KUENTZ-SIMONET, B. LIQUET, J. SARACCO. *ClustOfVar : an R package for dimension reduction via clustering of variables. Application in supervised classification and variable selection in gene expressions data*, in "Statistical Methods for (post)-Genomics Data (SMPGD 2013)", Netherlands, January 2013, http://hal.inria.fr/hal-00926216

[43] A. LABENNE, M. CHAVENT, V. KUENTZ-SIMONET, T. RAMBONILAZA, J. SARACCO. *Une approche par classification de variables à l'aide de la méthode 'ClustOfVar' pour analyser la qualité de vie à l'échelle communale*, in "Statlearn'13", Bordeaux, France, April 2013, http://hal.inria.fr/hal-00926229

[44] A. LABENNE, M. CHAVENT, V. KUENTZ-SIMONET, T. RAMBONILAZA, J. SARACCO. *Une extension de l'Analyse Factorielle Multiple pour des groupes de variables mixtes*, in "2èmes rencontres R", Lyon, France, June 2013, http://hal.inria.fr/hal-00926242

Scientific Books (or Scientific Book chapters)

[45] O. COSTA, F. DUFOUR. , *Continuous average control of piecewise deterministic Markov processes*, SpringerBriefs in Mathematics, Springer, 2013, 116 p. , http://hal.inria.fr/hal-00938693

[46] L. VEZARD, P. LEGRAND, M. CHAVENT, F. FAITA-AINSEBA, J. CLAUZEL, L. TRUJILLO. *Classification of EEG signals by evolutionary algorithm*, in "Advances in Knowledge Discovery and Management Volume 4", F. GUILLET, B. PINAUD, G. VENTURINI, D. ZIGHED (editors), Studies in Computational Intelligence, Springer, 2014, vol. 527, http://hal.inria.fr/hal-00939850

Research Reports

[47] H. ZHANG, B. DE SAPORTA, F. DUFOUR. , *Filtrage et contrôle optimal stochastique appliqué à l'optimisation de trajectoire*, May 2013, http://hal.inria.fr/hal-00926432

Other Publications

[48] R. COUDRET, S. GIRARD, J. SARACCO. , *A new sliced inverse regression method for multivariate response*, 2013, http://hal.inria.fr/hal-00714981

[49] A. GÉGOUT-PETIT, R. AZAÏS, M. TOUZET, M. PUIGGALI, A. BEN ABDESSALEM. , *Stochastic modelling and simulation of fatigue crack propagation using piecewise-deterministic Markov processes*, May 2013, Soumis, http://hal.inria.fr/hal-00940516

References in notes

[50] 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

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

[52] 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