RESEARCH CENTRE
Inria Nancy - Grand Est Center

IN PARTNERSHIP WITH:
CNRS, Université de Lorraine

2022
ACTIVITY REPORT

Project-Team
BIGS

Biology, genetics and statistics

IN COLLABORATION WITH: Institut Elie Cartan de Lorraine (IECL)

DOMAIN
Digital Health, Biology and Earth

THEME
Modeling and Control for Life Sciences

Inria

Contents

Project-Team BIGS

Creation of the Project-Team: 2011 January 01

Keywords

Computer sciences and digital sciences

A3.1. – Data

A3.2. – Knowledge

A3.2.3. – Inference

A3.3. – Data and knowledge analysis

A3.3.1. – On-line analytical processing

A3.3.2. – Data mining

A3.3.3. – Big data analysis

A3.4.1. – Supervised learning

A3.4.2. – Unsupervised learning

A3.4.4. – Optimization and learning

A3.4.7. – Kernel methods

A6. – Modeling, simulation and control

A6.1. – Methods in mathematical modeling

A6.1.2. – Stochastic Modeling

A6.2. – Scientific computing, Numerical Analysis & Optimization

A6.2.3. – Probabilistic methods

A6.2.4. – Statistical methods

A6.4. – Automatic control

A6.4.2. – Stochastic control

Other research topics and application domains

B1. – Life sciences

B1.1. – Biology

B1.1.2. – Molecular and cellular biology

B1.1.10. – Systems and synthetic biology

B1.1.11. – Plant Biology

B2.2. – Physiology and diseases

B2.2.1. – Cardiovascular and respiratory diseases

B2.2.3. – Cancer

B2.3. – Epidemiology

B2.4. – Therapies

1   Team members, visitors, external collaborators

Research Scientists

- Nicolas Champagnat [Team leader, INRIA, Senior Researcher, HDR]

- Coralie Fritsch [INRIA, Researcher]

- Ulysse Herbach [INRIA, Researcher]

- Bruno Scherrer [INRIA, Researcher, HDR]

Faculty Members

- Thierry Bastogne [UL, Associate Professor, HDR]

- Sandie Ferrigno [UL, Associate Professor]

- Anne Gégout-Petit [UL, Professor, HDR]

- Jean-Marie Monnez [UL, Emeritus, HDR]

- Aurélie Muller-Gueudin [UL, Associate Professor]

- Sophie Mézières [UL, Associate Professor]

- Pierre Vallois [UL, Emeritus, HDR]

- Denis Villemonais [UL, Associate Professor, HDR]

Post-Doctoral Fellows

- Léo Darrigade [INRIA, until Sep 2022]

- Joseph Lam-Weil [UL, until Jun 2022]

- William Oçafrain [INRIA, until Apr 2022]

PhD Students

- Virgile Brodu [ENS de Lyon, from Sep 2022]

- Vincent Hass [UL]

- Rodolphe Loubaton [UL]

- Anouk Rago [UL]

- Nino Vieillard [Google Brain, CIFRE, until Jun 2022]

- Nicolás Zalduendo Vidal [INRIA]

Technical Staff

- Walid Laziri [INRIA, Engineer, from Oct 2022]

Interns and Apprentices

- Virgile Brodu [ENS de Lyon, from Apr 2022 until Aug 2022]

Administrative Assistant

- Emmanuelle Deschamps [INRIA]

2   Overall objectives

BIGS is a joint team of Inria, CNRS and University of Lorraine, within the Institut Élie Cartan of Lorraine (IECL), UMR 7502 CNRS-UL laboratory in mathematics, of which Inria is a strong partner. One member of BIGS, T. Bastogne, comes from the Research Center of Automatic Control of Nancy (CRAN), with which BIGS has strong relations in the domain "Health-Biology-Signal". Our research is mainly focused on stochastic modeling and statistics but also aims at a better understanding of biological systems. BIGS involves applied mathematicians whose research interests mainly concern probability and statistics. More precisely, our attention is directed on (1) stochastic modeling, (2) estimation and control for stochastic processes, (3) regression and machine learning, and (4) statistical learning and application in health. The main objective of BIGS is to exploit these skills in applied mathematics to provide a better understanding of issues arising in life sciences, with a special focus on (1) tumor growth and heterogeneity, (2) gene networks, (3) telomere length dynamics, (4) epidemiology and e-health.

3   Research program

3.1   Introduction

We give here the main lines of our research. For clarity, we made the choice to structure them in four items. Note that all of these items deal with stochastic modeling and inference, therefore they are all interconnected.

3.2   Stochastic modeling

Our aim is to propose relevant stochastic frameworks for the modeling and the understanding of biological systems. The stochastic processes are particularly suitable for this purpose. Among them, Markov processes provide a first framework for the modeling of population of cells [85, 66]. Piecewise deterministic processes are non-diffusion processes that are also frequently used in the biological context [54, 65, 55]. Among Markov models, we developed strong expertise about processes derived from Brownian motion and Stochastic Differential Equations [80, 64]. For instance, knowledge about Brownian or random walk excursions [86, 79] helps to analyse genetic sequences and to develop inference about them. We also have strong expertise in stochastic modeling of complex biological populations using individual-based models. These models can be used either from the point of view of asymptotic stochastic analysis [51], e.g. to study the long term Darwinian evolution of populations, or from the point of view of numerical analysis of biological phenomena [60, 40]. We also develop mathematical tools for the analysis of the long-time behavior of stochastic population processes accounting for possible extinction of (sub)populations [52].

3.3   Estimation and control for stochastic processes

We develop inference about the stochastic processes that we use for modeling. Control of stochastic processes is also a way to optimise administration (dose, frequency) of therapy, such as targeted therapies in cancer. Our team has a good expertise about inference of the jump rate and the kernel of piecewise-deterministic Markov processes (PDMP) [45, 41, 44, 43], but there are many directions to go further into. For instance, previous work made the assumption of a complete observation of jumps and mode, which is unrealistic in practice. We also tackle the problem of inference of "hidden PDMP". For example, in pharmacokinetics modeling inference, we want to account for the presence of timing noise and identification from longitudinal data. We have expertise on these subjects [46], and we also use mixed models to estimate tumor growth or heterogeneity [47].

We consider the control of stochastic processes within the framework of Markov Decision Processes [77] and their generalization known as multi-player stochastic games, with a particular focus on infinite-horizon problems. In this context, we are interested in the complexity analysis of standard algorithms,

as well as the proposition and analysis of numerical approximate schemes for large problems in the spirit of [49]. Regarding complexity, a central topic of research is the analysis of the Policy Iteration algorithm, which has made significant progress in the last years [88, 76, 62, 82], but is still not fully understood. For large problems, we have an extensive experience of sensitivity analysis of approximate dynamic programming algorithms for Markov Decision Processes [81, 69, 83], and we currently investigate whether/how similar ideas may be adapted to multi-player stochastic games.

3.4   Algorithms and estimation for graph data

Recently, our group has focused its attention on modeling and inference for graph data. A graph data structure consists of a set of nodes, together with a set of pairs of these nodes called edges. This type of data is frequently used in biology because they provide a mathematical representation of many concepts such as biological networks of relationships in a population or between genes in a cell.

Network inference is the process of making inference about the link between two variables, by taking into account the information about other variables. [87] gives a very good introduction and many references about network inference and mining. Many methods are available to infer and test edges in Gaussian graphical models [87, 71, 59, 61]. However, the Gaussian assumption does not hold when dealing with typical "zero-inflated" abundance data, and we want to develop inference in this case.

Concerning gene networks, most studies have been based on population-averaged data: now that technologies enable us to observe mRNA levels in individual cells, a revolution in terms of precision, the network reconstruction problem paradoxically becomes more challenging than ever. Indeed, the traditional way of seeing a gene regulatory network as a deterministic system with some small external noise is being challenged by the probabilistic, bursty nature of gene expression revealed at single-cell level. Our objective is to propose dynamical models and inference methods that fully exploit the particular time structure of single-cell data. We described a promising strategy in which the network inference problem is seen as a calibration procedure for a new PDMP model that is able to acceptably reproduce real single-cell data [63, 78].

Among graphs, trees play a special role because they offer a powerful model for many biological concepts, from RNA to phylogenetic trees in heterogeneous tumors or through plant structures. Our research deals with several aspects of tree data. In particular, we work on statistical inference for this type of data under a given stochastic model. We also work on lossy compression of trees via directed acyclic graphs. These methods enable us to compute distances between tree data faster than from the original structures and with a high accuracy.

3.5   Regression and machine learning

Regression models and machine learning aim at inferring statistical links between a variable of interest and covariates. In biological studies, it is always important to develop adapted learning methods both in the context of *standard* data and also for data of high dimension (sometimes with few observations) and very massive or online data.

Many methods are available to estimate conditional quantiles and test dependencies [75, 67]. Among them we have developed nonparametric estimation by local analysis via kernel methods [57, 58] and we want to study properties of this estimator in order to derive a measure of risk based e.g. on confidence band and test. We study also other regression models like survival analysis, spatio-temporal models with covariates. Among the multiple regression models, we want to develop omnibus tests that examine several assumptions together.

Concerning the analysis of high dimensional data, our view on the topic relies on the *French data analysis school*, specifically on Factorial Analysis. In this context, stochastic approximation is an essential tool [68], which allows one to approximate eigenvectors in a stepwise manner [73, 72, 74]. We aim at performing accurate classification or clustering by taking advantage of the possibility of updating the information "online" using stochastic approximation algorithms [50]. We focus on several incremental procedures for regression and data analysis like linear and logistic regressions and PCA (Principal Component Analysis).

We also focus on the biological context of high-throughput bioassays in which several hundreds or thousands of biological signals are measured for a posterior analysis. We have to account for the

inter-individual variability within the modeling procedure. We aim at developing a new solution based on an ARX (Auto Regressive model with eXternal inputs) model structure using the EM (Expectation-Maximisation) algorithm for the estimation of the model parameters.

4 Application domains

4.1 Oncology: tumor growth and heterogeneity

We want to propose stochastic processes to model the appearance of mutations and the evolution of their frequencies in tumor samples, through new collaborations with clinicians who measure a particular quantity called circulating tumor DNA (ctDNA). The final purpose is to use ctDNA as an early biomarker of the resistance to a targeted therapy: this is the aim of the project funded by ITMO Cancer that we coordinate. In the ongoing work on low-grade gliomas, a local database of 400 patients will be soon available to construct models. We plan to extend it through national and international collaborations (Montpellier CHU, Montreal CRHUM). Our aim is to build a decision-aid tool for personalised medicine.

4.2 Gene networks and single-cell data

We already mentioned in Section 3.4 our interest in the modeling and inference of transcriptomic bursting in gene regulatory networks from single-cell data. We are also currently working on the prediction and identification of therapeutic targets for chronic lymphocytic leukemia from gene expression data. Our goal is to propose new models allowing to make prediction of gene silencing experiments. Inference will be performed on gene expression data from patients' cells suffering from different forms of chronic lymphocytic leukemia. The goal is to identify therapeutic targets which could be silenced to reduce cell proliferation.

4.3 Epidemiology and e-health

In the context of personalized medicine, we have many ongoing projects with CHU Nancy. They deal with biomarkers research, prognostic value of quantitative variables and events, scoring, and adverse events. We also want to develop our expertise in rupture detection in a project with APHP (Assistance Publique Hôpitaux de Paris) for the detection of adverse events, earlier than the clinical signs and symptoms. The clinical relevance of predictive analytics is obvious for high-risk patients such as those with solid organ transplantation or severe chronic respiratory disease for instance. The main challenge is the rupture detection in multivariate and heterogeneous signals (for instance daily measures of electrocardiogram, body temperature, spirometry parameters, sleep duration, etc.). Other collaborations with clinicians concern foetopathology and we want to use our work on conditional distribution function to explain fetal and child growth. To that end, we use data from the "Service de fœtopathologie et de placentologie" of the "Maternité Régionale Universitaire" (CHU Nancy).

4.4 Dynamics of telomeres

Telomeres are disposable buffers at the ends of chromosomes which are truncated during cell division; so that, over time, due to each cell division, the telomere ends become shorter. By this way, they are markers of aging. Through a collaboration with Pr A. Benetos, geriatrician at CHU Nancy, we recently obtained data on the distribution of the length of telomeres from blood cells [84]. We want to work in three connected directions: (1) refine methodology for the analysis of the available data; (2) propose a dynamical model for the lengths of telomeres and study its mathematical properties (long term behavior, quasi-stationarity, etc.); and (3) use these properties to develop new statistical methods.

5   New software and platforms

5.1   New software

5.1.1   quantCurves

Keyword: Statistical modeling

Functional Description: Non-parametric methods as local normal regression, polynomial local regression and penalized cubic B-splines regression are used to estimate quantiles curves.

News of the Year: Software developed in 2022.

URL: https://cran.r-project.org/web/packages/quantCurves/quantCurves.pdf

Contact: Sandie Ferrigno

5.1.2   Harissa

Name: Hartree approximation for inference along with a stochastic simulation algorithm

Keywords: Gene regulatory networks, Reverse engineering, Molecular simulation

Functional Description: Harissa is a Python package for both inference and simulation of gene regulatory networks, based on stochastic gene expression with transcriptional bursting. It was implemented in the context of a mechanistic approach to gene regulatory network inference from single-cell data.

News of the Year: New version in 2022 associated with an application paper (hal-03942224).

URL: https://github.com/ulysseherbach/harissa

Publications: hal-03942224, hal-03370296, hal-03370228, hal-01646910

Contact: Ulysse Herbach

6   New results

6.1   Stochastic modeling

Participants: Virgile Brodu, Nicolas Champagnat, Léo Darrigade, Coralie Fritsch, Anne Gégout-Petit, Vincent Hass, Ulysse Herbach, William Oçafrain, Pierre Vallois, Denis Villemonais, Nicolás Zalduendo Vidal.

6.1.1   Quasi-stationary distributions

We are continuing our research on quasi-stationary distributions (QSD), that is, distributions of Markov stochastic processes with absorption, which are stationary conditionally on non-absorption. For models of biological populations, absorption usually corresponds to extinction of a (sub-)population. QSDs are fundamental tools to describe the population state before extinction and to quantify the large-time behavior of the probability of extinction.

Thanks to the previous general result of the team in [53], together with B. Cloez (INRAE), we proved in [26] the exponential convergence of a chemostat model, whose dynamics are highly degenerate due to a deterministic part, towards a unique quasi-stationary distributions.

We also finalized an important work [7] that provides general criteria for the exponential convergence of conditional distributions of absorbed Markov processes when the convergence is not uniform with respect to the initial distribution. Our results allow to characterize a large subset of the domain of attraction of the minimal QSD and apply to a large range of stochastic processes, including diffusion

processes and perturbed dynamical systems. We completed this work with specific studies of the periodic case in [24] and the case of reducible processes in [25]. In this last work, we were in particular able to characterize cases with polynomial speed of convergence to a QSD and to prove the existence of a QSD for general processes in denumerable state spaces, assuming only aperiodicity, the existence of a Lyapunov function and the existence of a point in the state space from which the return time is finite with positive probability.

Motivated by our work with M. Benaïm (Univ. Neuchâtel) on degenerate processes such as hypoelliptic diffusions [48], we studied in [21] the links between Feller properties and quasi-compactness of general semigroups. This work allows to clarify the links between existing results on QSDs for hypoelliptic diffusions. We also provided in [6] a counterexample to the uniqueness of a quasi-stationary distribution for a diffusion process which satisfies the weak Hörmander condition.

In collaboration with A. Watson (Univ. College London, UK), we studied a general fragmentation-growth equation with unbounded rates. The work is decomposed into two parts: in the first part, it is shown that the equation admits a unique solution on a certain definition space; in the second part, spectral properties of the solution are established [36]. The proofs are largely based on [7] and on fine properties of piecewise deterministic Markov processes.

In [33], we obtained a central limit theorem and Berry-Esseen estimates for Markov processes conditioned to never be absorbed (so-called Q-processes) under the general conditions of [7]. We also obtained in [34] quasi-ergodic theorems for particular time-inhomogeneous Markov processes, whose time-inhomogeneity is asymptotically periodic, with applications to processes absorbed at a moving boundary.

6.1.2   Adaptive dynamics in biological populations

We continued our study of parameter scalings of individual-based models of biological populations under mutation and selection, taking into account the influence of negligible but non-extinct populations. In a work within the ERC SINGER, in collaboration with S. Méléard (École Polytechnique), S. Mirrahimi (Univ. Montpellier) and V.C. Tran (Univ. Paris Est Marne-la-Vallée) [23], we were able to give an individual-based justification of the Hamilton-Jacobi equation of adaptive dynamics (see e.g. [70]), with a specific parameter scaling that is promising for the study of local (in space) extinction of sub-populations. The analysis of models allowing for such an extinction is the next step of this project.

We also worked on general evolutionary models of adaptive dynamics under an assumption of large population and small mutations. This year, we obtained in [22] existence, uniqueness and ergodicity results for a centered version of the Fleming-Viot process of population genetics, which is a key step to recover variants of the canonical equation of adaptive dynamics, which describes the long time evolution of the dominant phenotype in the population, under less stringent biological assumptions than in previous works. We plan to complete this project next year.

6.1.3   Multi-type bisexual branching process and branching processes with Moran interaction

The asexual multi-type Galton-Watson branching processes as well as the single-type bisexual processes have been studied in the literature. In particular, survival condition of the processes are well known in both cases. However, until now, the multi-type bisexual branching processes have only been studied in very specific situations and no general mathematical description has been established yet.

In [31], we studied general multi-type bisexual branching processes with superadditive mating function. We exhibited a necessary and sufficient condition for almost sure extinction, we proved a law of large numbers for our model and we studied the long-time convergence of the rescaled process.

In collaboration with E. Horton (Inria ASTRAL team) and A. Cox (Univ. of Bath, UK), we proposed a new population size model gathering in the same framework branching processes and processes with Moran type interactions. We studied this process in a setting of large population and long time [27]. Its dynamics is related to the evolution of a non-conservative semi-group, whose spectral properties provide information on the long time behavior of the model.

6.1.4   Reconstruction of epigenetic landscapes from single-cell data

Joint work with E. Ventre (ENS Lyon), T. Espinasse (Univ. Lyon 1), G. Benoit (Univ. Rennes 1) and O.

Gandrillon (ENS Lyon).

The aim is to better understand how living cells make decisions (e.g., differentiation of a stem cell into a particular specialized type), seeing decision-making as an emergent property of an underlying complex molecular network. Indeed, it is now proven that cells react probabilistically to their environment: cell types do not correspond to fixed states, but rather to "potential wells" of a certain energy landscape (representing the energy of the possible states of the cell) that we are trying to reconstruct. The achievement of this year is to show that the same mathematical model driven by transcriptional bursting can be used simultaneously as an inference tool, to reconstruct biologically relevant networks, and as a simulation tool, to generate realistic transcriptional profiles emerging from gene interactions [35].

6.1.5   Modeling of chronic obstructive pulmonary disease

Joint work with I. Dupin (Univ. Bordeaux), E. Maurat (Univ. Bordeaux) and J.-M. Sac-Epée (IECL).

Lung exposure to various types of particules, such as those present in cigarette smoke, can lead to chronic obstructive pulmonary disease (COPD). COPD bronchi are an area of intense immunological activity and tissue remodeling, as evidenced by the extensive immune cell infiltration and changes in tissue structures. This allows the persistent contact between resident cells and stimulated immune cells. Our hypothesis is that the contact between cells is a major cause of chronic destructive or fibrotic manifestations. We aim to analyze the potential cell-cell interactions in situ in human tissues, to characterize in vitro the dynamics of the interplay, and to define a computational model with intercellular interactions which fits to experimental measurements and explains the macroscopic properties of cell populations. The effects of potential therapeutic drugs modulating local intercellular interactions will be tested by simulations. Two papers have been submitted this year [29, 30].

6.1.6   Numerical simulation of diffusions

In a collaboration with A. Lejay (Inria PASTA team) and their PhD student A. Anagnostakis, D. Villemonais proposed a method for approximating general, singular diffusions by discrete time and state space processes [1]. One of the main interests compared to existing methods is to propose a numerical method whose main computational cost is done upstream and thus represents a fixed cost, independently of the number of simulations performed afterwards.

6.1.7   A toy model of an animal foraging

We considered a toy model of an animal foraging in a one-dimensional space. The position of the animal as time t elapses is described by a standard Brownian motion. To survive, it needs one unit of food per unit of time, and it may stockpile any extra supply for future use, without any upper limit on the size of the stock nor any expiry date for its consumption. As for the provision of food, we assume that only half of the space is initially filled with one unit of food per unit length, and that there is no replenishment. We studied this model in [14], determining in particular the joint distribution of the position and time of death of the animal.

6.2   Regression and machine learning

Participants: Sandie Ferrigno, Jean-Marie Monnez.

6.2.1   Cramér–von Mises goodness-of-fit tests in regression models

Join work with R. Azaïs (Inria, ENS Lyon) and M.-J. Martinez (Univ. Grenoble Alpes).

Many goodness-of-fit tests have been developed to assess the different assumptions of a (possibly heteroscedastic) regression model. Most of them are 'directional' in that they detect departures from a given assumption of the model. Other tests are 'global' (or 'omnibus') in that they assess whether a model fits a dataset on all its assumptions. We focus on the task of choosing the structural part of the regression function because it contains easily interpretable information about the studied relationship. We consider two nonparametric 'directional' tests and one nonparametric 'global' test, all based on generalizations of the Cramér-von Mises statistic.

To perform these goodness-of-fit tests, we have developed the R package cvmgof [42], an easy-to-use tool for practitioners, available from the Comprehensive R Archive Network (CRAN). The use of the library is illustrated through a tutorial on real data and simulation studies are carried out in order to show how the package can be exploited to compare the 3 implemented tests. The practitioner can also easily compare the test procedures with different kernel functions, bootstrap distributions, numbers of bootstrap replicates, or bandwidths. The package was updated last year, this is its third version. An article [2] has been published on this work in 2022.

We are now working on nonparametric tests associated with the functional form of the variance of the regression model. For this, we continue to work on the global test of Ducharme and Ferrigno in order to compare its performance with directional tests associated with the variance of the model. Many simulations are in progress and a part of this work has been presented in CMStatistics 2022 conference [37]. This will also make it possible to propose a more general package-type tool allowing to validate the regression models used in practice.

To complete this work, we plan to assess the other assumptions of a regression model such as the additivity of the random error term. The implementation of these directional tests would enrich the cvmgof package and offer a complete easy-to-use tool for validating regression models. Moreover, the assessment of the overall validity of the model when using several directional tests will be compared with that done when using only a global test. In particular, we will discuss the well-known problem of multiple testing by comparing the results obtained from multiple test procedures with those obtained when using a global test strategy.

6.2.2   Online data analysis and online learning

Stochastic approximation is an important tool for the analysis of streaming data, introduced by Robbins and Monro in 1951, that can be used for example to estimate online parameters of a regression function [56] or centers of clusters in unsupervised classification [50]. Another type of stochastic approximation processes was introduced by Benzécri in 1969 for estimating eigenvectors and eigenvalues of the M-symmetric expectation of a random matrix A using independent observations of A. In all these processes, it is assumed that we can observe independent observations of the random matrix and that we take into account one or a mini-batch of observations per step. We are interested in the study of cases where we can't have independent observations and cases where at each step, we take into account all the observations up to this step without having to store them. Experiments we have conducted show that this second type of process generally converges faster than the first type with a mini-batch of observations at each step.

On this topic, our works with E. Albuisson (CHRU Nancy) on constrained binary logistic regression with online standardized data [13] and on the construction and update of an online ensemble score involving linear discriminant analysis and logistic regression [12], described in the previous activity report, are now published.

In the article [32], we establish an almost sure convergence theorem of two stochastic approximation processes for estimating eigenvectors of the unknown Q-symmetric expectation B of a random matrix: the first one is the classical process proposed by Oja and the second one makes use of past and current observations at each step. This theorem extends previous theorems to the case where the metric Q is unknown and estimated online in parallel. We apply these results to streaming PCA and streaming generalized canonical correlation analysis of a random vector Z, both when a mini-batch of observations of Z is used at each step or when all the observations up to the current step are used. Other applications are in progress, such as streaming multiple factor analyses and streaming partial PCA, as well as convergence theorems of Robbins-Monro type processes.

6.3 Statistical learning and application in health

Participants: Nicolas Champagnat, Sandie Ferrigno, Anne Gégout-Petit, Aurélie Gueudin, Walid Laziri, Rodolphe Loubaton, Sophie Mézières, Jean-Marie Monnez, Anouk Rago, Pierre Vallois.

6.3.1 Estimation of reference curves for fetal weight

Our research in the field of epidemiology focuses on fetal development in the last two trimesters of pregnancy. Reference or standard curves are required in this kind of biomedical problems. Values that lie outside the limits of these reference curves may indicate the presence of a disorder. Data are from the French EDEN mother-child cohort (INSERM). This is a mother-child cohort study investigating the prenatal and early postnatal determinants of child health and development. 2002 pregnant women were recruited before 24 weeks of amenorrhoea in two maternity clinics from middle-sized French cities (Nancy and Poitiers). From May 2003 to September 2006, 1899 newborns were then included. The main outcomes of interest are fetal (via ultra-sound) and postnatal growth, adiposity development, respiratory health, atopy, behaviour and bone, cognitive and motor development. We are studying fetal weight and height as a function of the gestional age in the third trimester of pregnancy. Some classical empirical and parametric methods such as polynomial regression are first used to construct these curves. For instance, polynomial regression is one of the most common parametric approaches for modeling growth data, especially during the prenatal period. However, these classical methods build upon restrictive assumptions on estimated curves. We therefore propose to work with semi-parametric LMS methods, by modifying the response variable (fetal weight) with, among others, Box–Cox transformations. An article detailing these methodologies applied to the EDEN data should be submitted next year.

Alternative nonparametric methods as Nadaraya-Watson kernel estimation, local polynomial estimation, B-splines or cubic splines are also developed to construct these curves. The practical implementation of these methods requires working on smoothing parameters or choice of knots for the different types of nonparametric estimation. In particular, optimal choice of these parameters has been proposed. To fit these curves, we have developed the R package quantCurves [39], an easy-to-use tool for practitioners. In addition, a graphical interface (GUI) intended for practitioners is being developed to enable intuitive visualization of the results given by the package and an article is in progress.

6.3.2 Prediction of silencing experiments on gene networks for chronic lymphocytic leukemia

We are working with L. Vallat (CHRU Strasbourg) on the inference of dynamical gene networks from RNAseq and proteome data. The goal is to infer a model of gene expression allowing to predict gene expression in cells where the expression of specific genes is silenced (e.g. using siRNA), in order to select the silencing experiments which are more likely to reduce the cell proliferation. We expect the selected genes to provide new therapeutic targets for the treatment of chronic lymphocytic leukemia. This year, we have developed a package for the statistical analysis of temporal gene expression datasets with several biological conditions (in particular for exploratory analysis and the detection of differentially expressed genes), which will be submitted soon to Bioconductor.

6.3.3 Covariates selection in high-dimensional data of genetic profiles in oncology

Joint work with T. Boukhobza and H. Dumond from CRAN, and B. Bastien from biopharmaceutical industry Transgene.

We proposed a new methodology for selecting and ranking covariates associated with a variable of interest in a context of high-dimensional data under dependence but few observations. The methodology successively intertwines the clustering of covariates, decorrelation of covariates using Factor Latent Analysis, selection using aggregation of adapted methods and finally ranking. We first applied our method to transcriptomic data of 37 patients with advanced non-small-cell lung cancer who have received chemotherapy, to select the transcriptomic covariates that explain the survival outcome of the treatment.

Secondly, we applied our method to 79 breast tumor samples to define patient profiles for a new metastatic biomarker and associated gene network in order to personalize the treatments. This work is published in [3] and is implemented in the R package 'ARMADA'.

6.3.4   Multidimensional statistical analysis of information for clinical use

The start-up EMOSIS develops blood tests relying on flow cytometry in order to improve in vitro diagnosis of vascular thrombosis. This technology leads to multiparametric measurements on tens of thousands cells collected from each blood sample. Manual methods of analysis classically used in flow cytometry are based on data visualization by means of histograms or scatter plots. Recent progresses in the active area of computational methods for dimension reduction suggest many directions of improvement of the classical approaches for the analysis of flow cytometry data. Our first goal is to define and operate such methods. We started to focus on methods of information geometry and topological analysis of data. Once appropriate methods will be identified, an important aspect to consider will be visualization of the results in a way easy to interpret by clinicians and ethically permissible. On a longer term, our ambition is to design more accurate prediction tools for diagnosis.

7   Bilateral contracts and grants with industry

Participants: Anne Gégout-Petit, Walid Laziri, Sophie Mézières, Bruno Scherrer.

7.1   Bilateral contracts with industry

- B. Scherrer collaborated with Google Brain on reinforcement learning in the framework of the PhD thesis of Nino Vieillard, until June 2022.

- As part of the French "Plan de relance", we obtained funds for a 2-year engineering contract with the start-up EMOSIS based in Strasbourg (from October 1, 2022). Project MOSAiC : MultidimensiOnal Statistical Analysis of Information for Clinical use.

8   Partnerships and cooperations

Participants: Virgile Brodu, Nicolas Champagnat, Léo Darrigade, Coralie Fritsch, Anne Gégout-Petit, Vincent Hass, Ulysse Herbach, Joseph Lam-Weil, Rodolphe Loubaton, Sophie Mézières, Jean-Marie Monnez, Aurélie Muller-Gueudin, Anouk Rago, Pierre Vallois, Denis Villemonais, Nicolás Zalduendo Vidal.

8.1   International initiatives

8.1.1   Associate Teams in the framework of an Inria International Lab or in the framework of an Inria International Program

MAGO

Title: Modelling and analysis for growth-fragmentation processes

Coordinator: E. Horton (Inria Bordeaux)

Partner Institutions: • Inria Bordeaux and Nancy (ASTRAL and BIGS teams, E. Horton, C. Fritsch, D. Villemonais)

- Univ. College of London (F.X. Briol, O. Key, A. Watson)

Date/Duration: 2022-2024

Total amount of the grant: 9000€ in 2022, in 2023 and 7000€ in 2024

Description: Growth-fragmentation (GF) refers to a collection of mathematical models in which objects – classically, biological cells – slowly gather mass over time, and fragment suddenly into multiple, smaller offspring. These models may be used to represent a range of biological processes, in which an individual reproduces by fission into two or more new individuals, such as the evolution of plasmids in bacteria populations and protein polymerisation. It is crucial to understand the long-term behaviour of GF processes so that they can be used to build algorithms to simulate real-world processes and estimate quantities such as the growth rate of the system, the steady state behaviour, and the fragmentation rate and kernel, allowing scientists to gain a better understanding of the behaviour of these complex systems. In this project, we aim to combine probabilistic and statistical tools to study these processes. In particular, we will employ methods from branching processes, quasi-stationary distributions and interacting particle systems to study their long-term behaviour and develop numerical simulations. Further, we will develop likelihood-free methods to estimate the model parameters, followed by goodness-of-fit tests to analyse the strength of these methods when working with real data.

8.2   European initiatives

8.2.1   ERC projects

N. Champagnat is scientific collaborator of the ERC SINGER (AdG 101054787) on Stochastic dynamics of sINgle cells, coordinated by S. Méléard (Ecole Polytechnique). He is involved in the research axes "From stochastic processes to singular Hamilton-Jacobi equations" and "Lineages and time reversed trajectories" of this project.

8.3   National initiatives

- FHU CARTAGE (Fédération Hospitalo Universitaire Cardial and ARTerial AGEing). **Leader:** Pr. A. Benetos. **Participants:** J.-M. Monnez, B. Lalloué, A. Gégout-Petit.

- RHU Fight HF (Fighting Heart Failure), located at the University Hospital of Nancy. **Leader:** Pr. P. Rossignol. **Participants:** J.-M. Monnez, B. Lalloué.

- ITMO Physics, Mathematics applied to Cancer (from 2017 until June 2022): "Modeling ctDNA dynamics for detecting targeted therapy resistance". **Funding organisms:** ITMO Cancer, ITMO Technologies pour la santé de l'alliance nationale pour les sciences de la vie et de la santé (AVIESAN), INCa. **Partners:** Inria and IECL (Institut Élie Cartan de Lorraine), CHRU Strasbourg, CRAN (Centre de Recherche en Automatique de Nancy) and ICL (Institut de Cancérologie de Lorraine). **Leader:** N. Champagnat. **Participants:** L. Darrigade, C. Fritsch, A. Gégout-Petit, U. Herbach, A. Muller-Gueudin, P. Vallois.

- GDR 720 ISIS (funded by CNRS). **Leader:** L. Blanc-Féraud. **Participant:** S. Mézières.

- Réseau Thématique MathSAV (funded by CNRS). **Leader:** Fabien Crauste. **Participants:** N. Champagnat, L. Darrigade, C. Fritsch, V. Hass, U. Herbach, J. Lam, R. Loubaton, A. Rago, N. Zalduendo Vidal.

- Chair "Modélisation Mathématique et Biodiversité" between VEOLIA, Ecole Polytechnique, Museum National d'Histoire Naturelle and Fondation X (funded by VEOLIA). **Leader:** S. Méléard. **Participants:** V. Brodu, N. Champagnat, C. Fritsch, V. Hass, D. Villemonais, N. Zalduendo Vidal.

8.4 Regional initiatives

- A regional fund PACTE has been obtained to host the Low-grade Glioma database and use it in diverse purposes: teaching, dissemination and development of experimentation tools. We continue to build the PIANO platform. **Participant:** S. Mézières.

- Région Grand-Est: in the context of the Telomere project, A. Gégout-Petit and Denis Villemonais obtained a grant from Grand-Est region to hire J. Lam-Weil as a post-doctoral fellow. University of Lorraine and LUE GEENAGE program completed the grant.

9 Dissemination

Participants:	Thierry Bastogne, Virgile Brodu, Nicolas Champagnat, Sandie Ferrigno, Coralie Fritsch, Anne Gégout-Petit, Vincent Hass, Ulysse Herbach, Rodolphe Loubaton, Sophie Mézières, Aurélie Muller-Gueudin, William Oçafrain, Anouk Rago, Bruno Scherrer, Denis Villemonais, Nicolás Zalduendo Vidal.

9.1 Promoting scientific activities

9.1.1 Scientific events: organization

Member of the organizing committees

- A. Gégout-Petit was organizer of the "Journées d'Etudes en Statistique" on compositional data, held in Fréjus in October.

- A. Gégout-Petit was co-organizer of the frENBIS event "Journées de Statistique industrielle", online, in April.

- U. Herbach is co-organizer of the Probability and Statistics weekly seminar at IECL in Nancy.

- D. Villemonais was organizer with Cécile Mailler (Univ. Bath) of the Workshop on Pólya urns, stochastic approximation and quasi-stationary distributions: new developments, held in Bath in April.

9.1.2 Scientific events: selection

Member of the conference program committees

- N. Champagnat was member of the scientific committee of JdS 2022 (53èmes Journées de Statistique de la SFdS), held in Lyon in June.

- In the framework of the Journées d'Etudes en Statistique 2021, A. Gégout-Petit was co-editor of a book on Missing data [19].

9.1.3 Journal

Member of the editorial boards N. Champagnat serves as associate editor of *ESAIM: Probability & Statistics* and *Stochastic Models*.

Reviewer - reviewing activities. The members of the team wrote referee reports for *Acta Applicandae Mathematicae, Annals of Applied Probability, Annals of Applied Statistics, Annals of Probability, Bulletin of Mathematical Biology, COVID, Discrete and Continuous Dynamical Systems Series B, Electronic Journal of Probability, Journal de l'École Polytechnique, Journal of Mathematical Biology, Stochastics and Partial Differential Equations: Analysis and Computations, Stochastic Models, Stochastic Processes and their Applications, Viruses.*

9.1.4 Invited talks

- N. Champagnat has been invited to give talks at the scientific day "Eco-Evo-Math: Building on Twenty Years of Research and Training at the Crossroads of Ecology, Evolution, and Mathematics" in Paris in July, at the "AG du département BioSiS du CRAN" in Nancy in May, at the Journées de Probabilités 2022 in Orbey in May, at the workshop on Pólya urns, stochastic approximation and quasi-stationary distributions in Bath in April, at the Journée cancérologie : innovations et expérience patient, d'hier à demain in Nancy in April, and at the workshop on Population Dynamics and Statistical Physics in Synergy in Oberwolfach in March.

- C. Fritsch has been invited to give talks at the Mathematical models in ecology and evolution in Paris in March, and at the Mathematical models in ecology and evolution conference in Reading in July.

- V. Hass has been invited to give talks at the Mathematical models in ecology and evolution conference in Reading in July, at MPDEE 2022 (Models in Population Dynamics, Ecology and Evolution) in Turin in June, and at the Journées de Probabilités 2022 in Orbey in May. He also gave seminar talks at the "Séminaire de l'équipe SPOC (Statistiques, Probabilités, Optimisation et Contrôle)" in Dijon in May, and at the "Séminaire de probabilités et systèmes dynamiques" in Amiens in March.

- U. Herbach has been invited to give talks at the London Institute for Mathematical Sciences in London in March, at the Laboratoire de Biologie et Modélisation de la Cellule in Lyon in April, at the Journées Math Bio Santé in Besançon in October, and at the Groupe de travail de Statistique in Rouen in December.

- W. Oçafrain has been invited to give a talk at the workshop on Pólya urns, stochastic approximation and quasi-stationary distributions in Bath in April.

- D. Villemonais has been invited to give talks at the Workshop "Singular diffusions: numerical and theoretical aspects" in Nancy in May, at the Journées de Probabilités 2022 in Orbey in May and at the workshop on Pólya urns, stochastic approximation and quasi-stationary distributions in Bath in April. He also gave seminar talks at the Séminaire de calcul stochastique de Strasbourg in December in Strasbourg, at the Séminaire de probabilités d'Orsay in November in Orsay and at the Séminaire FHU - CARTAGE in Nancy in November.

- N. Zalduendo Vidal has been invited to give talks at the Journées de Probabilités 2022 in Orbey in May, and at the Mathematical models in ecology and evolution conference in Reading in July.

9.1.5 Leadership within the scientific community

A. Gégout-Petit is vice-president of the European Network for Business and Industrial Statistics (ENBIS).

9.1.6 Scientific expertise

A. Gégout-Petit was in the hiring committees of a biostatistics 'Professeur' position at Sorbonne Univ., and three 'Chaire de professeur junior' (CPJ) at CNRS (INSMI, SPJ Monaie), Univ. Lorraine (Biostatistics), and Univ. de Pau et des Pays de l'Adour (Artificial Intelligence).

9.1.7 Research administration

- N. Champagnat is a member of the COMIPERS and the Commission Information Scientifique et Technique of Inria Nancy - Grand Est and Responsable Scientifique for the library of Mathematics of the IECL. He is also local correspondent of the COERLE (Comité Opérationel d'Évaluation des Risques Légaux et Éthiques) for the Inria Research Center of Nancy - Grand Est.

- C. Fritsch is a member of the Commission du Développement Technologique of Inria Nancy-Grand Est and of the Commission du personnel of IECL. She was the local Radar correspondent for the Inria Research Center of Nancy - Grand Est until June.

- A. Gégout-Petit is the head of IECL.

9.2 Teaching - Supervision - Juries

9.2.1 Teaching

BIGS faculty members have teaching obligations at Univ. Lorraine and are teaching at least 192 hours each year. They teach probability and statistics at different levels (Licence, Master, Engineering school). Many of them have pedagogical responsibilities.

- D. Villemonais is the head of the Mathematical Engineering Major of ENSMN, Université de Lorraine.

- T. Bastogne is in charge of the research master program "Santé Numérique et Imagerie Médicale" with the Faculty of Medicine, Université de Lorraine.

- Licence: V. Brodu, Probability Theory tutorial, 40h, L3, first year of ENSMN, Université de Lorraine.

- Licence: V. Brodu, Numerical Analysis tutorial, 20h, L3, first year of ENSMN, Université de Lorraine.

- Master: N. Champagnat, Introduction to Quantitative Finance, 12h, M1, second year of ENSMN, Université de Lorraine.

- Master: N. Champagnat, Introduction to Quantitative Finance, 9h, M2, third year of ENSMN, Université de Lorraine.

- Master: S. Ferrigno, Experimental designs, 4.5h, M1, fourth year of EEIGM, Université de Lorraine.

- Master: S. Ferrigno, Data analyzing and mining, 36h, M1, second year of ENSMN, Université de Lorraine.

- Master: S. Ferrigno, Modeling and forecasting, 32h, M1, second year of ENSMN, Université de Lorraine.

- Master: S. Ferrigno, Training projects, 18h, M1/M2, second and third year of ENSMN, Université de Lorraine.

- Licence: S. Ferrigno, Descriptive and inferential statistics, 60h, L2, second year of EEIGM, Université de Lorraine.

- Licence: S. Ferrigno, Statistical modeling, 60h, L2, second year of EEIGM, Université de Lorraine.

- Licence: S. Ferrigno, Mathematical and computational tools, 20h, L3, third year of EEIGM, Université de Lorraine.

- Licence: S. Ferrigno, Training projects, 40h, L1/L3, first, second and third year of EEIGM, Université de Lorraine.

- Master: C. Fritsch, Inverse problem, 18h, M1, second year of ENSMN, Université de Lorraine.

- Licence: C. Fritsch, Probability Theory tutorial, 40h, L3, first year of ENSMN, Université de Lorraine.

- Master: A. Gégout-Petit, Statistics, modeling, data analysis, 80h, master in applied mathematics, Université de Lorraine.

- Licence: V. Hass, Mathématiques FIGIM 1A, 38h, L1/L2, first year of ENSMN, Université de Lorraine.

- Licence: V. Hass, Mathématiques FIGIM 2A, 19h, L2, second year of ENSMN, Université de Lorraine.

- Licence: V. Hass, Probabilités, 40h, L3, first year of ENSMN, Université de Lorraine.

- Licence: V. Hass, Analyse numérique et optimisation, 45h, L3, first year of ENSMN, Université de Lorraine.

- Licence: V. Hass, Recherche opérationnelle, 18h, L3, first year of ENSMN, Université de Lorraine.

- Master: V. Hass, Méthodes stochastiques pour le calcul, 14h, M1, second year of ENSMN, Université de Lorraine.

- Licence: V. Hass, Mathématiques FIGIM 1A, 70h, L1/L2, first year of ENSMN, Université de Lorraine.

- Licence: V. Hass, Mathématiques FIGIM 2A, 19h, L2, second year of ENSMN, Université de Lorraine.

- Licence: V. Hass, Probabilités, 40h, L3, first year of ENSMN, Université de Lorraine.

- Licence: V. Hass, Analyse numérique et optimisation, 45h, L3, first year of ENSMN, Université de Lorraine.

- Licence: V. Hass, Recherche opérationnelle, 18h, L3, first year of ENSMN, Université de Lorraine.

- Licence: R. Loubaton, Inférence statistique, 42h, L3, first year of ENSMN, Université de Lorraine.

- Master: R. Loubaton, Analyse de données, 18h, M1, second year of ENSMN, Université de Lorraine.

- Master: R. Loubaton, Introduction à l'apprentissage automatique, 14h, M1, second year of ENSMN, Université de Lorraine.

- Master: R. Loubaton, Introduction au deep learning, 14h, M1, second year of ENSMN, Université de Lorraine.

- Licence: R. Loubaton, Analyse numérique, 44h, L3, first year of ENSMN, Université de Lorraine.

- Licence: R. Loubaton, Remédiation mathématique pour étudiants étrangers, 36h, L3, first year of ENSMN, Université de Lorraine.

- Licence: R. Loubaton, Géométrie et vecteurs pour la physique, 25h, L1, first year of EEIGM, Université de Lorraine.

- Licence: R. Loubaton, Analyse, 25h, L1, first year of ENGSI, Université de Lorraine.

- Master : A. Rago, Modélisation et Prévision, 14h, M1, second year of ENSMN, Université de Lorraine.

- Licence : A. Rago, Analyse numérique et optimisation, 20h, L3, first year of ENSMN, Université de Lorraine.

- Master : A. Rago, Analyse de données, 18h, M1, second year of ENSMN, Université de Lorraine.

- Master : A. Rago, Statistiques pour la grande dimension, 18h, M2 IMSD/third year of ENSMN, Université de Lorraine.

- Master: D. Villemonais, Probability Theory II, 63h, M1, second year of ENSMN, Université de Lorraine.

- Master: D. Villemonais, Stochastic processes, 32h, Master 2 MFA, Université de Lorraine.

- Master: D. Villemonais, Modeling and forecasting, 14h, M1, second year of ENSMN, Université de Lorraine.

- License: D. Villemonais Probability Theory I, 57h, L3, first year of ENSMN, Université de Lorraine.

- Master: S. Wantz-Mézières, Learning and analysis of medical data, 36h, with J.M. Moureaux, M2 SNIM, Université de Lorraine.

- Licence: S. Wantz-Mézières, Applied mathematics for management, financial mathematics, Probability and Statistics, 160h, IUT Nancy-Charlemagne (L1/L2/L3), Université de Lorraine.

- Licence: S. Wantz-Mézières, Probability, 100h, first year in TELECOM Nancy (initial and apprenticeship cursus), Université de Lorraine.

- Licence: N. Zalduendo Vidal, Probability Theory tutorial, 40h, L3, first year of ENSMN, Université de Lorraine.

- Licence: N. Zalduendo Vidal, Numerical Analysis tutorial, 20h, L3, first year of ENSMN, Université de Lorraine.

9.2.2   Supervision

PhD

- PhD in progress: Virgile Brodu, "Émergence des allométries dans les systèmes écologiques : comportement stationnaire de modèles déterministes et stochastiques de flux d'énergie et de biomasse", grant ENS Lyon. Advisors: S. Billiard (Univ. Lille), N. Champagnat, C. Fritsch.

- PhD in progress: Vincent Hass, "Individual-based models in adaptive dynamics and long time evolution under assumptions of rare advantageous mutations", grant Inria-Cordi, currently ATER in ENSMN. Advisor: N. Champagnat.

- PhD in progress: Rodolphe Loubaton, "Caractérisation des cibles thérapeutiques dans un programme génique tumoral", grant Région Grand-Est, currently ATER in EEIGM. Advisors: N. Champagnat and L. Vallat (CHRU Strasbourg).

- PhD in progress: Anouk Rago, "Inférence de réseaux de gènes dynamiques et prédiction d'expériences d'interventions biologiques dans des cellules cancéreuses", grant Région Grand-Est, Inria. Advisors: N. Champagnat, A. Gégout-Petit.

- PhD: Nino Vieillard, "Approximate Dynamic Programming and Deep Reinforcement Learning", CIFRE with Google Brain. Advisors: B. Scherrer, M. Geist (Google Brain), defense on June, 30.

- PhD in progress: Nicolás Zalduendo Vidal, "Processus de branchement bi-sexués multi-types", grant Inria-Cordis. Advisors: C. Fritsch, D. Villemonais.

Other

- M2 internship: Virgile Brodu, "Emergence des allométries dans les systèmes écologiques : convergence d'un modèle individu-centré de flux d'énergie vers la solution d'un système d'équations intégro-différentielles" (ENS Lyon). Advisor: N. Champagnat, C. Fritsch and S. Billiard (Univ. Lille).

- Research project: Hassan Berrada, "Condition de survie et d'extinction pour un modèle bisexué" (M2 ENSMN). Advisor: C. Fritsch, N. Zalduendo Vidal.

- Research project: Hugo Breton, "Interface graphique et package quantCurves" (M2 ENSMN). Advisor : S. Ferrigno.

- Parcours Recherche: Romain Maillard, "Inférence statistique de réseaux de gènes à partir de graphes dynamiques" (full-year research project, M1 ENSMN). Advisor: U. Herbach.

- Project M1: Guillaume Nodet et May Ouir, "Inférence de réseaux de gènes avec Random Forest" (M1 ENSMN). Advisor: R. Loubaton.

9.2.3   Juries

- PhD: N. Champagnat, reviewer, PhD thesis of Apolline Louvet, "Modèles probabilistes de génétique des populations pour les populations en expansion", Institut polytechnique de Paris.

- PhD: A. Gégout-Petit, reviewer, PhD thesis of Guillaume Bottaz-Bottom, "Classification de trajectoires d'observances de patients atteints d'un syndrome d'apnées obstructives du sommeil", Univ. Grenoble-Alpes.

- HDR: A. Gégout-Petit, reviewer, HDR thesis of Frédéric Proia, "Autorégressifs à coefficients variables — Modèles graphiques partiels — Applications aux sciences du vivant", Univ. Angers.

- HDR: A. Gégout-Petit, HDR thesis of Romain Azaïs, "Approches algorithmiques pour la statistique : processus déterministes par morceaux et arbres aléatoires", ENS Lyon.

- PhD: A. Gégout-Petit, thesis of Olivier Coudray, "Un point de vue statistique sur les critères de fatigue: de la classification supervisée à l'apprentissage positif-non labellisé", Université Paris-Saclay.

- PhD: A. Gégout-Petit, thesis of Cécile Spychala, "Statistical analysis of road accidents in the region Franche-Comté: risk factors for accident injuries and spatial modelling for accident occurrences", Univ. Besançon, Franche-Comté.

- PhD: B. Scherrer, reviewer, PhD thesis of Léonard Blier, "Some Principled Methodes for Deep Reinforcement Learning", Univ. Paris-Saclay.

- PhD: B. Scherrer, reviewer, PhD thesis of Giovanni Gatti Pinheiro, "Apprentissage par renforcement appliqué au Revenue Management des compagnies aériennes", Univ. Côte d'Azur.

- PhD: B. Scherrer, reviewer, PhD thesis of Chen Yan, "Asymptotically Optimal Policies for Restless Bandits", Univ. Grenoble-Alpes.

- Prize: A. Gégout-Petit, member of the committe for AMIES PhD prize.

9.3   Popularization

9.3.1   Education

- S. Mézières: organisation of a research training week on NeuroOncology and Numerics, for medical and engineering students, January 2022.

9.3.2   Interventions

- C. Fritsch made two interventions in the Lycée Cormontaigne in Metz, as part of the "Chiche!" program, in November.

- S. Ferrigno: Advisor of a group of students, "Traitement statistique de données" Project, various high schools, Nancy.

- S. Ferrigno: Advisor of a group of students, "La main à la Pâte" Project, Institut médico-éducatif (IME), Commercy.

- S. Ferrigno: Advisor of a group of students, "La main à la Pâte", "C'Génial" Projects, Colleges, Malzéville and Nancy.

- S. Ferrigno: Advisor of a group of students, "La main à la Pâte" Project, elementary schools, Nancy.

- U. Herbach gave a general public conference "Les maths peuvent-elles servir à vaincre le cancer ?" in Ambert in October, for the breast cancer national awareness campaign "Octobre Rose". On this occasion, he also made several scientific mediation interventions in secondary and high schools of Ambert.

- R. Loubaton gave an introduction to artificial intelligence in the conference "Être humain à l'âge de l'IA" at Paris in July.

- J-M. Monnez gave a masterclass on online data analysis [38] to the BIGS working group in Nancy, in October and November.

10   Scientific production

10.1   Publications of the year

International journals

[1]   A. Anagnostakis, A. Lejay and D. Villemonais. 'General diffusion processes as the limit of time-space Markov chains'. In: *Annals of Applied Probability* (2023). URL: https://hal.inria.fr/hal-028 97819.

[2]   R. Azaïs, S. Ferrigno and M.-J. Martinez. 'cvmgof: an R package for Cramér-von Mises goodness-of-fit tests in regression models'. In: *Journal of Statistical Computation and Simulation* 92.6 (2022), pp. 1246–1266. DOI: 10.1080/00949655.2021.1991346. URL: https://hal.science/hal-03101612.

[3]   B. Bastien, T. Boukhobza, H. Dumond, A. Gégout-Petit, A. Muller-Gueudin and C. Thiébaut. 'A statistical methodology to select covariates in high-dimensional data under dependence. Application to the classification of genetic profiles in oncology'. In: *Journal of Applied Statistics* 49.3 (Mar. 2022), pp. 764–781. DOI: 10.1080/02664763.2020.1837083. URL: https://hal.science/hal-02173568.

[4]   T. Bastogne. 'iQbD: a TRL-indexed quality-by-design paradigm for medical device engineering'. In: *Journal of Medical Devices* 16.2 (June 2022), p. 021008. DOI: 10.1115/1.4053721. URL: https://hal.archives-ouvertes.fr/hal-03710686.

[5]   T. Bastogne, F. Caputo, A. Prina-Mello, S. Borgos and M. Barberi-Heyob. 'A state of the art in analytical quality-by-design and perspectives in characterization of nano-enabled medicinal products'. In: *Journal of Pharmaceutical and Biomedical Analysis* 219 (22nd Sept. 2022), p. 114911. DOI: 10.1016/j.jpba.2022.114911. URL: https://hal.archives-ouvertes.fr/hal-03710680.

[6]   M. Benaïm, N. Champagnat, W. Oçafrain and D. Villemonais. 'Transcritical bifurcation for the conditional distribution of a diffusion process'. In: *Journal of Theoretical Probability* (28th Nov. 2022). DOI: 10.1007/s10959-022-01216-7. URL: https://hal.science/hal-03471911.

[7]   N. Champagnat and D. Villemonais. 'General criteria for the study of quasi-stationarity'. In: *Electronic Journal of Probability* (2023). URL: https://hal.inria.fr/hal-01672979.

[8]   L. Darrigade, M. Haghebaert, C. Cherbuy, S. Labarthe and B. Laroche. 'A PDMP model of the epithelial cell turn-over in the intestinal crypt including microbiota-derived regulations'. In: *Journal of Mathematical Biology* 84.7 (June 2022), pp. 1–67. DOI: 10.1007/s00285-022-01766-8. URL: https://hal.science/hal-03379023.

[9]   S. Deneuve, T. Bastogne, M. Duclos, C. Mirjolet, P. Bois, P. Bachmann, L. Nokovitch, P.-E. Roux, D. Girodet, M. Poupart, P. Zrounba, L. Claude, L. Ferella, A. Iacovelli, N. Foray, T. Rancati and S. Pereira. 'Predicting acute severe toxicity for head and neck squamous cell carcinomas by combining dosimetry with a radiosensitivity biomarker : a pilot study'. In: *Tumori* (May 2022). DOI: 10.1177/03008916221078061. URL: https://hal.archives-ouvertes.fr/hal-03718454.

[10]   A. Dijamentiuk, C. Mangavel, A. Elfassy, F. Michaux, J. Burgain, E. Rondags, S. Delaunay, S. Ferrigno, A.-M. Revol-Junelles and F. Borges. 'Invert emulsions alleviate biotic interactions in bacterial mixed culture'. In: *Microbial Cell Factories* 22.1 (Dec. 2023), p. 16. DOI: 10.1186/s12934-022-02014-w. URL: https://hal.univ-lorraine.fr/hal-03953726.

[11]   M. Geoffroy, M. Lemesle, A. Kleinclauss, S. Mazerbourg, L. Batista, M. Barberi-Heyob, T. Bastogne, W. Boireau, A. Rouleau, D. Dupommier, M. Boisbrun, C. Comoy, S. Flament, I. Grillier-Vuissoz and S. Kuntz. 'AB186 inhibits migration of triple-negative breast cancer cells and interacts with α-Tubulin'. In: *International Journal of Molecular Sciences* 23.12 (June 2022), p. 6859. DOI: 10.3390/ijms23126859. URL: https://hal.archives-ouvertes.fr/hal-03705488.

[12]   B. Lalloué, J.-M. Monnez and E. Albuisson. 'Construction and Update of an Online Ensemble Score Involving Linear Discriminant Analysis and Logistic Regression'. In: *Applied Mathematics* 13.2 (Feb. 2022), pp. 228–242. DOI: 10.4236/am.2022.132018. URL: https://hal.archives-ouvertes.fr/hal-03134248.

[13]   B. Lalloué, J.-M. Monnez and E. Albuisson. 'Streaming constrained binary logistic regression with online standardized data'. In: *Journal of Applied Statistics* 49.6 (2022), pp. 1519–1539. DOI: 10.1080/02664763.2020.1870672. URL: https://hal.science/hal-02156324.

[14]   J. Randon-Furling, P. Salminen and P. Vallois. 'On a first hit distribution of the running maximum of Brownian motion'. In: *Stochastic Processes and their Applications* 150 (15th June 2022). DOI: 10.1016/j.spa.2021.12.015. URL: https://hal.science/hal-03170250.

[15] L. Schenone, C. Houillier, M. L. Tanguy, S. Choquet, K. Agbetiafa, H. Ghesquières, G. Damaj, A. Schmitt, K. Bouabdallah, G. Ahle, R. Gressin, J. Cornillon, R. Houot, J.-P. Marolleau, L.-M. Fornecker, O. Chinot, F. Peyrade, R. Bouabdallah, C. Moluçon-Chabrot, E. Gyan, A. Chauchet, O. Casasnovas, L. Oberic, V. Delwail, J. Abraham, V. Roland, A. Waultier-Rascalou, L. Willems, F. Morschhauser, M. Fabbro, R. Ursu, C. Thieblemont, F. Jardin, A. Tempescul, D. Malaise, V. Touitou, L. Nichelli, M. Le Garff-Tavernier, A. Plessier, P. Bourget, C. Bonmati, S. Wantz-Mézières, Q. Giordan, V. Dorvaux, C. Charron, W. Jabeur, K. Hoang-Xuan, L. Taillandier and C. Soussain. 'Intensive chemotherapy followed by autologous stem cell transplantation in primary central nervous system lymphomas (PCNSLs). Therapeutic outcomes in real life-experience of the French network'. In: *Bone Marrow Transplantation* 57.6 (Apr. 2022), pp. 966–974. DOI: 10.1038/s41409-022-01648-z. URL: https://hal-u-picardie.archives-ouvertes.fr/hal-03703797.

[16] A. Taccoen, C. Piedallu, I. Seynave, A. Gégout-Petit and J.-C. Gégout. 'Climate change-induced background tree mortality is exacerbated towards the warm limits of the species ranges'. In: *Annals of Forest Science* 79.1 (Dec. 2022), p. 23. DOI: 10.1186/s13595-022-01142-y. URL: https://hal.science/hal-03694879.

International peer-reviewed conferences

[17] S. Rezaeifar, R. Dadashi, N. Vieillard, L. Hussenot, O. Bachem, O. Pietquin and M. Geist. 'Offline Reinforcement Learning as Anti-Exploration'. In: AAAI 2022 - 36th AAAI Conference on Artificial Intelligence. Vancouver, Canada, 22nd Feb. 2022. URL: https://hal.science/hal-03468875.

Scientific books

[18] F. Bertrand, M. Maumy-Bertrand, A. Muller, S. Ferrigno, D. Marx and Y. Rabba Idi. *Mathématiques pour les sciences de l'ingénieur - Tout le cours en fiches: 3ème édition.* Dunod, 2022, 576 pages. URL: https://hal.archives-ouvertes.fr/hal-03576344.

[19] A. Gégout-Petit, M. Maumy-Bertrand, G. Saporta and C. Thomas-Agnan. *Données manquantes.* Editions Technip, 16th June 2022. URL: https://hal-cnam.archives-ouvertes.fr/hal-03696270.

Scientific book chapters

[20] A. Gégout-Petit, M. Maumy-Bertrand, G. Saporta and C. Thomas-Agnan. 'Une histoire lacunaire'. In: *Données manquantes*. Editions Technip, June 2022, pp. 1–27. URL: https://hal-cnam.archives-ouvertes.fr/hal-03696274.

Reports & preprints

[21] M. Benaïm, N. Champagnat, W. Oçafrain and D. Villemonais. *Quasi-compactness criterion for strong Feller kernels with an application to quasi-stationary distributions.* 13th Apr. 2022. URL: https://hal.archives-ouvertes.fr/hal-03640205.

[22] N. Champagnat and V. Hass. *Existence, uniqueness and ergodicity for the centered Fleming-Viot process.* 3rd Mar. 2022. URL: https://hal.archives-ouvertes.fr/hal-03596777.

[23] N. Champagnat, S. Méléard, S. Mirrahimi and V.-C. Tran. *Filling the gap between individual-based evolutionary models and Hamilton-Jacobi equations.* 2nd May 2022. URL: https://hal.archives-ouvertes.fr/hal-03656608.

[24] N. Champagnat and D. Villemonais. *Quasi-limiting estimates for periodic absorbed Markov chains.* 2022. URL: https://hal.inria.fr/hal-03823655.

[25] N. Champagnat and D. Villemonais. *Quasi-stationary distributions in reducible state spaces.* 24th Jan. 2022. URL: https://hal.archives-ouvertes.fr/hal-03541248.

[26] B. Cloez and C. Fritsch. *Quasi-stationary behavior for an hybrid model of chemostat: the Crump-Young model.* 9th May 2022. URL: https://hal.inria.fr/hal-03660678.

[27]   A. M. Cox, E. Horton and D. Villemonais. *Binary branching processes with Moran type interactions.*
        7th July 2022. URL: https://hal.inria.fr/hal-03715678.

[28]   A. Deveau, A. Gégout-Petit and C. Karmann. *Penalized polytomous ordinal logistic regression using
        cumulative logits. Application to network inference of zero-inflated variables.* 22nd Aug. 2022. URL:
        https://hal.inrae.fr/hal-03757473.

[29]   I. Dupin, E. Eyraud, É. Maurat, J.-M. Sac-Epee and P. Vallois. *Probabilistic Cellular Automata
        modeling of intercellular interactions in airways : complex pattern formation in patients with
        Chronic Obstructive Pulmonary Disease.* 12th Oct. 2022. URL: https://hal.science/hal-03572
        045.

[30]   E. Eyraud, E. Maurat, J.-M. Sac-Epee, P. Henrot, M. Zysman, P. Esteves, T. Trian, H. Bégueret, P.-O.
        Girodet, M. Thumerel, R. Hustache-Castaing, R. Marthan, F. Levet, P. Vallois, C. Contin-Bordes,
        P. Berger and I. Dupin. *Short-range interactions between fibrocytes and CD8+ T cells in COPD
        bronchial inflammatory response.* 21st Oct. 2022. DOI: 10.1101/2022.10.21.513138. URL:
        https://hal.science/hal-03958836.

[31]   C. Fritsch, D. Villemonais and N. Zalduendo. *The Multi-type Bisexual Galton-Watson Branching
        Process.* 15th June 2022. URL: https://hal.archives-ouvertes.fr/hal-03696115.

[32]   J.-M. Monnez. *Stochastic approximation of eigenvectors and eigenvalues of the Q -symmetric expec-
        tation of a random matrix.* 2022. DOI: 10.1080/03610926.2022.2107225. URL: https://hal.s
        cience/hal-03956687.

[33]   W. Oçafrain. *A central limit and Berry-Esseen theorem for continuous-time Markov processes condi-
        tioned not to be absorbed.* 6th Mar. 2022. URL: https://hal.archives-ouvertes.fr/hal-0359
        9148.

[34]   W. Oçafrain. *An ergodic theorem for asymptotically periodic time-inhomogeneous Markov processes,
        with application to quasi-stationarity with moving boundaries.* 5th Apr. 2022. URL: https://hal
        .archives-ouvertes.fr/hal-02963683.

[35]   E. Ventre, U. Herbach, T. Espinasse, G. Benoit and O. Gandrillon. *One model fits all: combining
        inference and simulation of gene regulatory networks.* 20th June 2022. DOI: 10.1101/2022.06.19
        .496754. URL: https://hal.inria.fr/hal-03942224.

[36]   D. Villemonais and A. Watson. *A quasi-stationary approach to the long-term asymptotics of the
        growth-fragmentation equation.* 24th Feb. 2022. URL: https://hal.archives-ouvertes.fr/ha
        l-03588374.

Other scientific publications

[37]   S. Ferrigno and M.-J. Martinez. 'GOODNESS-OF-FIT TESTS FOR VARIANCE FUNCTION IN RE-
        GRESSION MODELS'. In: CMStatistics 2022. London, United Kingdom, 17th Dec. 2022. URL: https
        ://hal.science/hal-03935703.

10.2   Other

Educational activities

[38]   J.-M. Monnez. 'Analyse des données en flux. Analyse en composantes principales et méthodes
        dérivées'. Doctoral. France, 18th Oct. 2022. URL: https://hal.archives-ouvertes.fr/hal-0
        3690459.

Softwares

[39]   [SW] S. Ferrigno, D. Essaket, A. Mouchot, H. Breton and M. Maumy-Bertrand, *quantCurves:Estimate
        Quantiles Curves* version 1.0.0, 11th Mar. 2022. LIC: CeCILL. HAL: ⟨hal-03683008⟩, URL: https:
        //hal.science/hal-03683008, SWHID: ⟨swh:1:dir:0dada269f87de4a5764bc7ba37e9c584
        3c273e34;origin=https://hal.archives-ouvertes.fr/hal-03683008;visit=swh:1:sn
        p:9d0ab5466bba609329cceb209dcf50d4067585e8;anchor=swh:1:rel:28715855bcbd3f48
        6c88f9982e40a71d26f42d9d;path=/⟩.

10.3   Cited publications

[40]   M. Andrade-Restrepo, N. Champagnat and R. Ferrière. 'Spatial eco-evolutionary dynamics along environmental gradients: multi-stability and cluster dynamics'. In: *Ecology Letters* 22.5 (May 2019), pp. 767–777. URL: https://hal.inria.fr/hal-01732325.

[41]   R. Azaïs, F. Dufour and A. Gégout-Petit. 'Non-Parametric Estimation of the Conditional Distribution of the Interjumping Times for Piecewise-Deterministic Markov Processes'. In: *Scandinavian Journal of Statistics* 41.4 (Dec. 2014), pp. 950–969. DOI: 10.1111/sjos.12076. URL: https://hal.archives-ouvertes.fr/hal-01103700.

[42]   [SW] R. Azaïs, S. Ferrigno and M.-J. Martinez, *cvmgof: Cramer-von Mises goodness-of-fit tests* version 1.0.0, Nov. 2018. LIC: CeCILL. HAL: ⟨hal-02014516v1⟩, URL: https://hal.archives-ouvertes.fr/hal-02014516, SWHID: ⟨swh:1:dir:6b9e75ac30a78bdf2d928b2177e8edf538c16748⟩.

[43]   R. Azaïs and A. Muller-Gueudin. 'Optimal choice among a class of nonparametric estimators of the jump rate for piecewise-deterministic Markov processes'. In: *Electronic journal of statistics* (2016). URL: https://hal.archives-ouvertes.fr/hal-01168651.

[44]   R. Azaïs. 'A recursive nonparametric estimator for the transition kernel of a piecewise-deterministic Markov process'. In: *ESAIM: Probability and Statistics* 18 (2014), pp. 726–749.

[45]   R. Azaïs, F. Dufour and A. Gégout-Petit. 'Nonparametric estimation of the jump rate for non-homogeneous marked renewal processes'. In: *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques*. Vol. 49. 4. Institut Henri Poincaré. 2013, pp. 1204–1231.

[46]   T. Bastogne, S. Mézières-Wantz, N. Ramdani, P. Vallois and M. Barberi-Heyob. 'Identification of pharmacokinetics models in the presence of timing noise'. In: *Eur. J. Control* 14.2 (2008), pp. 149–157. DOI: 10.3166/ejc.14.149-157. URL: http://dx.doi.org/10.3166/ejc.14.149-157.

[47]   T. Bastogne, A. Samson, P. Vallois, S. Wantz-Mézières, S. Pinel, D. Bechet and M. Barberi-Heyob. 'Phenomenological modeling of tumor diameter growth based on a mixed effects model'. In: *Journal of theoretical biology* 262.3 (2010), pp. 544–552.

[48]   M. Benaïm, N. Champagnat, W. Oçafrain and D. Villemonais. 'Degenerate processes killed at the boundary of a domain'. working paper or preprint. 2021. URL: https://hal.inria.fr/hal-03278551.

[49]   D. Bertsekas and J. Tsitsiklis. *Neurodynamic Programming*. Athena Scientific, 1996.

[50]   H. Cardot, P. Cénac and J.-M. Monnez. 'A fast and recursive algorithm for clustering large datasets with k-medians'. In: *Computational Statistics and Data Analysis* 56 (2012), pp. 1434–1449. DOI: 10.1016/j.csda.2011.11.019. URL: https://hal.science/hal-00644683.

[51]   N. Champagnat, P.-E. Jabin and S. Méléard. 'Adaptation in a stochastic multi-resources chemostat model'. In: *Journal de Mathématiques Pures et Appliquées* 101.6 (June 2014), pp. 755–788. DOI: 10.1016/j.matpur.2013.10.003. URL: https://hal.inria.fr/hal-00784166.

[52]   N. Champagnat and D. Villemonais. 'Exponential convergence to quasi-stationary distribution and Q-process'. In: *Probability Theory and Related Fields* 164.1 (2016). 46 pages, pp. 243–283. DOI: 10.1007/s00440-014-0611-7. URL: https://hal.archives-ouvertes.fr/hal-00973509.

[53]   N. Champagnat and D. Villemonais. 'Practical criteria for R-positive recurrence of unbounded semigroups'. In: *Electronic Communications in Probability* 25.none (2020), pp. 1–11. DOI: 10.1214/20-ECP288. URL: https://doi.org/10.1214/20-ECP288.

[54]   M. H. Davis. 'Piecewise-deterministic Markov processes: A general class of non-diffusion stochastic models'. In: *Journal of the Royal Statistical Society. Series B (Methodological)* (1984), pp. 353–388.

[55]   M. Doumic, M. Hoffmann, N. Krell and L. Robert. 'Statistical estimation of a growth-fragmentation model observed on a genealogical tree'. In: *Bernoulli* 21.3 (2015), pp. 1760–1799.

[56]   K. Duarte, J.-M. Monnez and E. Albuisson. 'Sequential linear regression with online standardized data'. In: *PLoS ONE* (2018), pp. 1–27. DOI: 10.1371/journal.pone.0191186. URL: https://hal.science/hal-01538125.

[57] S. Ferrigno and G. Ducharme. 'Un test d'adéquation global pour la fonction de répartition conditionnelle'. In: *C. R. Math. Acad. Sci. Paris* 341.5 (2005), pp. 313–316. DOI: 10.1016/j.crma.2005.07.003. URL: http://dx.doi.org/10.1016/j.crma.2005.07.003.

[58] S. Ferrigno, M. Maumy-Bertrand and A. Muller-Gueudin. 'Uniform law of the logarithm for the local linear estimator of the conditional distribution function'. In: *C. R. Math. Acad. Sci. Paris* 348.17-18 (2010), pp. 1015–1019. DOI: 10.1016/j.crma.2010.08.003. URL: http://dx.doi.org/10.1016/j.crma.2010.08.003.

[59] J. Friedman, T. Hastie and R. Tibshirani. 'Sparse inverse covariance estimation with the graphical lasso'. In: *Biostatistics* 9.3 (2008), pp. 432–441.

[60] C. Fritsch, F. Campillo and O. Ovaskainen. 'A numerical approach to determine mutant invasion fitness and evolutionary singular strategies'. In: *Theoretical Population Biology* 115 (2017), pp. 89–99. DOI: 10.1016/j.tpb.2017.05.001. URL: https://hal.science/hal-01413638.

[61] C. Giraud, S. Huet and N. Verzelen. 'Graph selection with GGMselect'. In: *Statistical applications in genetics and molecular biology* 11.3 (2012).

[62] T. Hansen and U. Zwick. 'Lower Bounds for Howard's Algorithm for Finding Minimum Mean-Cost Cycles'. In: *ISAAC (1)*. 2010, pp. 415–426.

[63] U. Herbach, A. Bonnaffoux, T. Espinasse and O. Gandrillon. 'Inferring gene regulatory networks from single-cell data: a mechanistic approach'. In: *BMC Systems Biology* 11.1 (Nov. 2017), p. 105. DOI: 10.1186/s12918-017-0487-0. URL: https://hal.science/hal-01646910.

[64] S. Herrmann and P. Vallois. 'From persistent random walk to the telegraph noise'. In: *Stoch. Dyn.* 10.2 (2010), pp. 161–196. DOI: 10.1142/S0219493710002905. URL: http://dx.doi.org/10.1142/S0219493710002905.

[65] J. Hu, W.-C. Wu and S. Sastry. 'Modeling subtilin production in bacillus subtilis using stochastic hybrid systems'. In: *Hybrid Systems: Computation and Control*. Springer, 2004, pp. 417–431.

[66] R. Keinj, T. Bastogne and P. Vallois. 'Multinomial model-based formulations of TCP and NTCP for radiotherapy treatment planning'. In: *Journal of Theoretical Biology* 279.1 (June 2011), pp. 55–62. DOI: 10.1016/j.jtbi.2011.03.025. URL: http://hal.inria.fr/hal-00588935/en.

[67] R. Koenker. *Quantile regression*. 38. Cambridge university press, 2005.

[68] L. Lebart. 'On the Benzecri's method for computing eigenvectors by stochastic approximation (the case of binary data)'. In: *Compstat 1974 (Proc. Sympos. Computational Statist., Univ. Vienna, Vienna, 1974)*. Vienna: Physica Verlag, 1974, pp. 202–211.

[69] B. Lesner and B. Scherrer. 'Non-Stationary Approximate Modified Policy Iteration'. In: *ICML 2015*. Lille, France, July 2015. URL: https://hal.inria.fr/hal-01186664.

[70] A. Lorz, S. Mirrahimi and B. Perthame. 'Dirac mass dynamics in multidimensional nonlocal parabolic equations'. In: *Communications in Partial Differential Equations* 36.6 (2011), pp. 1071–1098.

[71] N. Meinshausen and P. Bühlmann. 'High-dimensional graphs and variable selection with the lasso'. In: *The Annals of Statistics* (2006), pp. 1436–1462.

[72] J.-M. Monnez. 'Approximation stochastique en analyse factorielle multiple'. In: *Ann. I.S.U.P.* 50.3 (2006), pp. 27–45.

[73] J.-M. Monnez. 'Convergence d'un processus d'approximation stochastique en analyse factorielle'. In: *Publ. Inst. Statist. Univ. Paris* 38.1 (1994), pp. 37–55.

[74] J.-M. Monnez. 'Stochastic approximation of the factors of a generalized canonical correlation analysis'. In: *Statist. Probab. Lett.* 78.14 (2008), pp. 2210–2216. DOI: 10.1016/j.spl.2008.01.088. URL: http://dx.doi.org/10.1016/j.spl.2008.01.088.

[75] E. Nadaraya. 'On non-parametric estimates of density functions and regression curves'. In: *Theory of Probability & Its Applications* 10.1 (1965), pp. 186–190.

[76] I. Post and Y. Ye. *The simplex method is strongly polynomial for deterministic Markov decision processes*. Tech. rep. arXiv:1208.5083v2, 2012.

[77]   M. Puterman. *Markov Decision Processes*. Wiley, New York, 1994.

[78]   A. Richard, L. Boullu, U. Herbach, A. Arnaud, V. Morin, E. Vallin, A. Guillemin, N. Papili Gao, R. Gunawan, J. Cosette, O. Arnaud, J.-J. Kupiec, T. Espinasse, S. Gonin-Giraud, O. Gandrillon and S. Teichmann. 'Single-Cell-Based Analysis Highlights a Surge in Cell-to-Cell Molecular Variability Preceding Irreversible Commitment in a Differentiation Process'. In: *PLoS Biology* 14.12 (Dec. 2016). DOI: 10.1371/journal.pbio.1002585. URL: https://hal.science/hal-01934489.

[79]   B. Roynette, P. Vallois and M. Yor. 'Brownian penalisations related to excursion lengths, VII'. In: *Annales de l'IHP Probabilités et statistiques*. Vol. 45. 2. 2009, pp. 421–452.

[80]   F. Russo and P. Vallois. 'Stochastic calculus with respect to continuous finite quadratic variation processes'. In: *Stochastics: An International Journal of Probability and Stochastic Processes* 70.1-2 (2000), pp. 1–40.

[81]   B. Scherrer. 'Approximate Policy Iteration Schemes: A Comparison'. In: *ICML - 31st International Conference on Machine Learning - 2014*. Pékin, China, June 2014. URL: https://hal.inria.fr/hal-00989982.

[82]   B. Scherrer. 'Improved and Generalized Upper Bounds on the Complexity of Policy Iteration'. In: *Mathematics of Operations Research* (Feb. 2016). Markov decision processes ; Dynamic Programming ; Analysis of Algorithms. DOI: 10.1287/moor.2015.0753. URL: https://hal.inria.fr/hal-00829532.

[83]   B. Scherrer, M. Ghavamzadeh, V. Gabillon, B. Lesner and M. Geist. 'Approximate Modified Policy Iteration and its Application to the Game of Tetris'. In: *Journal of Machine Learning Research* 16 (2015). A paraître, pp. 1629–1676. URL: https://hal.inria.fr/hal-01091341.

[84]   S. Toupance, D. Villemonais, D. Germain, A. Gégout-Petit, E. Albuisson and A. Benetos. 'The individual's signature of telomere length distribution'. In: *Scientific Reports* 9.1 (Jan. 2019), pp. 1–8. DOI: 10.1038/s41598-018-36756-8. URL: https://hal.inria.fr/hal-01925000.

[85]   P. Vallois. 'The range of a simple random walk on Z'. In: *Advances in applied probability* (1996), pp. 1014–1033.

[86]   P. Vallois and C. S. Tapiero. 'Memory-based persistence in a counting random walk process'. In: *Phys. A.* 386.1 (2007), pp. 303–307. DOI: 10.1016/j.physa.2007.08.027. URL: http://dx.doi.org/10.1016/j.physa.2007.08.027.

[87]   N. Villa-Vialaneix. *An introduction to network inference and mining*. http://wikistat.fr/. (consulté le 22/07/2015). 2015. URL: http://www.nathalievilla.org/doc/pdf//wikistat-network_compiled.pdf.

[88]   Y. Ye. 'The Simplex and Policy-Iteration Methods Are Strongly Polynomial for the Markov Decision Problem with a Fixed Discount Rate'. In: *Math. Oper. Res.* 36.4 (2011), pp. 593–603.