Activity Report 2011

Project-Team CLASSIC

Computational Learning, Aggregation, Supervised Statistical, Inference, and Classification

IN COLLABORATION WITH: Département de Mathématiques et Applications (DMA)

Table of contents

Project-Team CLASSIC

Keywords: Machine Learning, Statistical Learning, Sequential Learning, Game Theory, Classification, Information Theory

The team has been created on July the 1st, 2009 and became an INRIA project on January the 1st, 2010.

1. Members

Research Scientists
Olivier Catoni [Team leader, Senior researcher, CNRS, HdR]
Gilles Stoltz [Junior researcher, CNRS, HdR]

Faculty Member
Gérard Biau [Professor, Université Paris Pierre-et-Marie-Curie, HdR]

External Collaborator
Vincent Rivoirard [Professor, Université Paris-Dauphine, HdR]

PhD Students
Sébastien Gerchinovitz [PhD student, fellow of Université Paris-Sud]
Thomas Mainguy [PhD student, fellow of Ecole normale supérieure]
Emilien Joly [PhD student, student at ENS de Cachan]
Pierre Gaillard [PhD student, student at Ecole normale supérieure]

Post-Doctoral Fellow
Jia Yuan Yu [Post-doctoral fellow, funded by the ANR project EXPLO/RA]

Administrative Assistants
Hélène Milome [Assistant till November 2011; shared with other teams]
Marie-Noëlle Julienne [Assistant since November 2011; shared with other teams]

2. Overall Objectives

2.1. Overall Objectives

We are a research team on machine learning, with an emphasis on statistical methods. Processing huge amounts of complex data has created a need for statistical methods which could remain valid under very weak hypotheses, in very high dimensional spaces. Our aim is to contribute to a robust, adaptive, computationally efficient and desirably non-asymptotic theory of statistics which could be profitable to learning.

Our theoretical studies bear on the following mathematical tools:

– regression models used for supervised learning, from different perspectives: the PAC-Bayesian approach to generalization bounds; robust estimators; model selection and model aggregation;

– sparse models of prediction and ℓ_1–regularization;

– interactions between unsupervised learning, information theory and adaptive data representation;

– individual sequence theory;

– multi-armed bandit problems (possibly indexed by a continuous set).

We are involved in the following applications:

– improving prediction through the on-line aggregation of predictors applied to air quality control, electricity consumption, stock management in the retail supply chain;
– natural image analysis, and more precisely the use of unsupervised learning in data representation;
– computational linguistics;
– statistical inference on biological data.

3. Scientific Foundations

3.1. Regression models of supervised learning

The most obvious contribution of statistics to machine learning is to consider the supervised learning scenario as a special case of regression estimation: given n independent pairs of observations (X_i, Y_i), $i = 1, \cdots, n$, the aim is to "learn" the dependence of Y_i on X_i. Thus, classical results about statistical regression estimation apply, with the caveat that the hypotheses we can reasonably assume about the distribution of the pairs (X_i, Y_i) are much weaker than what is usually considered in statistical studies. The aim here is to assume very little, maybe only independence of the observed sequence of input-output pairs, and to validate model and variable selection schemes. These schemes should produce the best possible approximation of the joint distribution of (X_i, Y_i) within some restricted family of models. Their performance is evaluated according to some measure of discrepancy between distributions, a standard choice being to use the Kullback-Leibler divergence.

3.1.1. PAC-Bayes inequalities

One of the specialties of the team in this direction is to use PAC-Bayes inequalities to combine thresholded exponential moment inequalities. The name of this theory comes from its founder, David McAllester, and may be misleading. Indeed, its cornerstone is rather made of non-asymptotic entropy inequalities, and a perturbative approach to parameter estimation. The team has made major contributions to the theory, first focussed on classification [6], then on regression [1]. It has introduced the idea of combining the PAC-Bayesian approach with the use of thresholded exponential moments, in order to derive bounds under very weak assumptions on the noise.

3.1.2. Sparsity and ℓ_1–regularization

Another line of research in regression estimation is the use of sparse models, and its link with ℓ_1–regularization. Regularization is the joint minimization of some empirical criterion and some penalty function; it should lead to a model that not only fits well the data but is also as simple as possible.

For instance, the Lasso uses a ℓ^1–regularization instead of a ℓ^0–one; it is popular mostly because it leads to *sparse* solutions (the estimate has only a few nonzero coordinates), which usually have a clear interpretation in many settings (e.g., the influence or lack of influence of some variables). In addition, unlike ℓ^0–penalization, the Lasso is *computationally feasible* for high-dimensional data.

3.1.3. Pushing it to the extreme: no assumption on the data

The next brick of our scientific foundations explains why and how, in certains cases, we may formulate absolutely no assumption on the data (x_i, y_i), $i = 1, \cdots, n$, which is then considered a deterministic set of input–output pairs.

3.2. On-line aggregation of predictors for the prediction of time series, with or without stationarity assumptions

We are concerned here with *sequential prediction* of outcomes, given some base predictions formed by *experts*. We distinguish two settings, depending on how the sequence of outcomes is generated: it is either

– the realization of some stationary process,
– or is not modeled at all as the realization of any underlying stochastic process (these sequences are called *individual sequences*).

The aim is to predict almost as well as the best expert. Typical good forecasters maintain one weight per expert, update these weights depending on the past performances, and output at each step the corresponding weighted linear combination of experts' advices.

The difference between the cumulative prediction error of the forecaster and the one of the best expert is called the regret. The game consists here of upper bounding the regret by a quantity as small as possible.

3.3. Multi-armed bandit problems, prediction with limited feedback

We are interested here in settings in which the feedback obtained on the predictions is limited, in the sense that it does not fully reveal what actually happened.

3.3.1. Bandit problems

This is also a sequential problem in which some regret is to be minimized.

However, this problem is a stochastic problem: a large number of arms, possibly indexed by a continuous set like $[0, 1]$, is available. Each arm is associated with a fixed but unknown distribution. At each round, the player chooses an arm, a payoff is drawn at random according to the distribution that is associated with it, and the only feedback that the player gets is the value of this payoff. The key quantity to study this problem is the mean-payoff function f, that indicates for each arm x the expected payoff $f(x)$ of the distribution that is associated with it. The target is to minimize the regret, i.e., ensure that the difference between the cumulative payoff obtained by the player and the one of the best arm is small.

3.3.2. A generalization of the regret: the approachability of sets

Approachability is the ability to control random walks. At each round, a vector payoff is obtained by the first player, depending on his action and on the action of the opponent player. The aim is to ensure that the average of the vector payoffs converges to some convex set. Necessary and sufficient conditions were obtained by Blackwell and others to ensure that such strategies exist, both in the full information and in the bandit cases.

Some of these results can be extended to the case of games with signals (games with partial monitoring), where at each round the only feedback obtained by the first player is a random signal drawn according to a distribution that depends on the action profile taken by the two players, while the opponent player still has a full monitoring.

4. Application Domains

4.1. Forecasting of the electricity consumption

Our partner is EDF R&D. The goal is to aggregate in a sequential fashion the forecasts made by some (about 20) base experts in order to predict the electricity consumption at a global level (the one of all French customers) at a half-hourly step. We need to abide by some operational constraints: the predictions need to be made at noon for the next 24 hours (i.e., for the next 48 time rounds).

4.2. Forecasting of the air quality

Our partner is the INRIA project-team CLIME (Paris-Rocquencourt). The goal is to aggregate in a sequential fashion the forecasts made by some (about 100) base experts in order to output field prediction of the concentration of some pollutants (typically, the ozone) over Europe. The results were and will be transferred to the public operator INERIS, which uses and will use them in an operational way.

4.3. Management of the supply chain

Our partner is the start-up Lokad.com. The purpose of this application is to investigate nonparametric expert-oriented strategies for time series prediction from a practical perspective.

4.4. Computational linguistics

The aim is to propose and study new language models that bridge the gap between models oriented towards statistical analysis of large corpora and grammars oriented towards the description of syntactic features as understood by academic experts.

4.5. Statistical inference on biological data

We have here two specific applications in mind.

One is about understanding how the transcription of human genome is performed: transcription regulatory elements need to be identified. A natural modeling is provided by multivariate Hawkes processes but an excessive computational time is necessary for their implementation. Lasso type methods may overcome this numerical issue.

The second is about estimating the division rate of a size-structured population in a nonparametric setting. The size of the system evolves according to a transport-fragmentation equation: each individual grows with a given transport rate and splits into two offsprings of the same size, following a binary fragmentation process with an unknown division rate that depends on its size.

5. New Results

5.1. Contributions earlier to 2011 but only published in 2011

Participants: Gérard Biau, Olivier Catoni, Sébastien Gerchinovitz, Vincent Rivoirard, Gilles Stoltz.

We do not discuss here the contributions provided by [12], [13], [14], [17], [18], [20], [22], [23], [25], since they were achieved in 2009 or earlier (but only published this year due to long queues in publication tracks of journals). [32] was revised but is still under review.

5.2. Sparse regression estimation

Participants: Gérard Biau, Olivier Catoni, Sébastien Gerchinovitz, Vincent Rivoirard, Gilles Stoltz, Jia Yuan Yu.

Sébastien Gerchinovitz and Jia Yuan Yu continued the work initiated by the former in the above-mentioned conference paper [25]; they derived from the sparsity results in individual sequences presented therein the minimax optimal rates of aggregation for individual sequences on ℓ^1 balls. In particular, they exhibited, in certain cases, a phase transition between the $\ln T$ and the $\sqrt{T}$ behavior of the minimax regret, where T denotes the number of instances. These results and all previous ones are summarized in the PhD thesis [10].

Other results were obtained in a stochastic framework, where input–output pairs are given by i.i.d. variables; they are described in the technical report [30]. Let $(\mathbf{X}, Y)$ be a random pair taking values in $\mathbb{R}^p \times \mathbb{R}$. In the so-called single-index model, one has $Y = f^*(\theta^{*T}\mathbf{X}) + \mathbf{W}$, where f^* is an unknown univariate measurable function, θ^* is an unknown vector in $\mathbb{R}^d$, and W denotes a random noise satisfying $\mathbb{E}[\mathbf{W}|\mathbf{X}] = 0$. The single-index model is known to offer a flexible way to model a variety of high-dimensional real-world phenomena. However, despite its relative simplicity, this dimension reduction scheme faces severe complications as soon as the underlying dimension becomes larger than the number of observations and this is why this estimation problem was considered from a sparsity perspective using a PAC-Bayesian approach.

Last but not least, we mention the edited book [29], which provides a modern overview on high-dimensional estimation.

5.3. Sequential learning with limited feedback; in particular, bandit problems

Participants: Gilles Stoltz, Jia Yuan Yu.

Some of the results cited below are summarized or stated as open problems in the habilitation thesis [11].

5.3.1. *Bandit problems*

We achieved three contributions. The first is described in the conference paper [27]: it revisits asymptotically optimal results of Lai and Robbins, Burnetas and Katehakis in a non-asymptotic way. The second is stated in the journal article [19] and is concerned with obtaining fast convergence rates for the regret in case of a continuum of arms (of course under some regularity and topological assumptions on the mean-payoff function f).

The third one is detailed in [24] and started from the following observation. Typical results in the bandit literature were of the following form: if the regularity of the mean-payoff function f is known (or if a bound on it is known) then the regret is small. Actually, results were usually taking the following weaker form: when the algorithm is tuned with some parameters, then the regret is small against a certain class of stochastic environments. The question was thus to have an adaptive procedure, that, given one unknown environment (with unknown regularity), ensures that the regret is asymptotically small; even better, the desired aim was to control the regret in some uniform manner (in a distribution-free sense up to the regularity parameters). As described in this conference paper, a solution was achieved in the case of Lipschitz environments.

5.3.2. *Approachability in games with partial monitoring*

The conference paper [28] explains how we could re-obtain, in a simple, more straightforward, and computationally efficient manner a result proven by Perchet in his PhD thesis: the necessary and sufficient condition for the approachability of a closed convex set under partial monitoring.

5.4. Inference

Participant: Gérard Biau.

5.4.1. *Geometric inference*

This line of research is in collaboration with the Geometrica project-team (INRIA Saclay). As the latter says:

> Due to the fast evolution of data acquisition devices and computational power, scientists in many areas are demanding efficient algorithmic tools for analyzing, manipulating and visualizing more and more complex shapes or complex systems from approximating data. Many of the existing algorithmic solutions which come with little theoretical guarantees provide unsatisfactory and/or unpredictable results. Since these algorithms take as input discrete geometric data, it is mandatory to develop concepts that are rich enough to robustly and correctly approximate continuous shapes and their geometric properties by discrete models. Ensuring the correctness of geometric estimations and approximations on discrete data is a sensitive problem in many applications.

Thus, motivated by a broad range of potential applications in topological and geometric inference, we introduce in [15] a weighted version of the k–nearest neighbor density estimator. Various pointwise consistency results of this estimator are established; the proposed method is also implemented to recover level sets in both simulated and real-life data.

Another problem of geometric inference is the following one, studied in [16]. Principal curves are nonlinear generalizations of the notion of first principal component. Roughly, a principal curve is a parameterized curve in $\mathbb{R}^d$ that passes through the "middle" of a data cloud drawn from some unknown probability distribution. Depending on the definition, a principal curve relies on some unknown parameters (number of segments, length, turn...) which have to be properly chosen to recover the shape of the data without interpolating. In this paper, we consider the principal curve problem from an empirical risk minimization perspective and address the parameter selection issue using the point of view of model selection via penalization. We offer oracle inequalities and implement the proposed approaches to recover the hidden structures in both simulated and real-life data.

5.4.2. Statistical inference

We still keep an eye on more traditional mathematical statistics; in particular, the technical report [31] takes place within this field. It shows, for a large class of distributions and large samples, that estimates of the variance σ^2 and of the standard deviation σ are more often Pitman closer to their target than the corresponding shrinkage estimates which improve the mean squared error. The results thus indicate that Pitman closeness criterion, despite its controversial nature, should be regarded as a useful and complementary tool for the evaluation of estimates of σ^2 and of σ.

5.5. Statistical inference for biological systems based on a size-structured population

Participant: Vincent Rivoirard.

The journal paper [21] considers the problem of estimating the division rate of a size-structured population in a nonparametric setting. The size of the system evolves according to a transport-fragmentation equation: each individual grows with a given transport rate, and splits into two offsprings of the same size, following a binary fragmentation process with unknown division rate that depends on its size. In contrast to a deterministic inverse problem approach, this paper takes the perspective of statistical inference: the data consists in a large sample of the size of individuals when the evolution of the system is close to its time-asymptotic behavior, so that it can be related to the eigenproblem of the considered transport-fragmentation equation. By estimating statistically each term of the eigenvalue problem and suitably inverting a certain linear operator, it constructs a more realistic estimator of the division rate that achieves the same optimal error bound as in related deterministic inverse problems. The procedure relies on kernel methods with automatic bandwidth selection. It is inspired by model selection and recent results of Goldenschluger and Lepski.

6. Contracts and Grants with Industry

6.1. Contracts with Industry

Gérard Biau has been supervising the PhD thesis of Benoît Patra, which takes places within an industrial contract ("thèse CIFRE") with Lokad.com (http://www.lokad.com/).

7. Partnerships and Cooperations

7.1. National Initiatives

We (co-)organized the following seminars:

– Statistical machine learning in Paris – SMILE (Gérard Biau, Gilles Stoltz; see http://sites.google.com/site/smileinparis/);

– Parisian seminar of statistics at IHP (Vincent Rivoirard; see https://sites.google.com/site/semstats).

Grants:

– ANR project in the conception and simulation track: EXPLO/RA (involves Emilien Joly, Sébastien Gerchinovitz, Gilles Stoltz, Jia Yuan Yu; see http://sites.google.com/site/anrexplora/);

– ANR project in the blank program: Parcimonie (involves Sébastien Gerchinovitz, Vincent Rivoirard, Gilles Stoltz; see http://www.proba.jussieu.fr/ANR/Parcimonie/);

– two other ANR blank projects only involve each one member of the team: Banhdits (Vincent Rivoirard), CLARA (Gérard Biau).

7.2. European Initiatives

Thanks to the PASCAL European network of Excellence (http://www.pascal-network.org/), we have strong links with Gábor Lugosi, Universitat Pompeu Fabra, Spain and Nicolò Cesa-Bianchi, Università degli Studi di Milano.

7.3. International Initiatives

We have some internal collaborations, mostly on one-to-one bases, with

– Karine Bertin, University of Valparaiso, Chile;

– Luc Devroye, McGill University, Canada;

– Shie Mannor, Technion, Israel.

8. Dissemination

8.1. Animation of the scientific community

8.1.1. *Editorial activities, reports written on articles*

Gérard Biau serves as an Associate Editor for the journals *Annales de l'ISUP*, *ESAIM: Probability and Statistics* and *International Statistical Review*.

Olivier Catoni is a member of the editorial committee of the joint series of monographies "Mathématiques et Applications" between Springer and SMAI.

All permanent members of the team reviewed several journal papers during the year.

8.1.2. *Report written on PhD / habilitation theses, participation to defense committees*

We wrote reports on PhD (1 by Gilles Stoltz) and habilitation (1 by Olivier Catoni) theses.

We were examinators for other PhD (4 by Gérard Biau, 1 by Olivier Catoni) and habilitation (1 by Gérard Biau) defenses.

8.1.3. *Participation to national or local evaluation or recruitment committees, to scientific societies*

Vincent Rivoirard was elected at the Conseil de la SFdS.

Gérard Biau was elected member of the national council of French universities (CNU) within the applied mathematics section (number 26).

Olivier Catoni is a member of the doctoral commission in mathematics of Universities Pierre et Marie Curie and Paris Diderot.

All permanent members of the team participated in several recruitment committees for assistant or full professors in universities.

8.1.4. *Conference organization*

Gilles Stoltz was a member of the program committee of the 24th Conference on Learning Theory (COLT'11); Vincent Rivoirard was a member of the program committee of the Journées de la SFdS 2011.

Gérard Biau organized with Pierre Alquier in December 2011 an invited session entitled "High-dimensional statistics, sparsity and applications" at the 4th International Conference of the ERCIM Working Group on Computing & Statistics.

8.1.5. Dissemination of scientific knowledge to the general audience

Gille Stoltz participated in a meeting [Rencontres S'Cube] between a crowd of 4 professional mathematicians and a general audience; the theme was "Perdre ou gagner, peut-on prévoir ?" and the meeting took place in Gif-sur-Yvette, in May 2011.

8.2. Teaching

The permanent members of the team (Gérard Biau, Olivier Catoni, Vincent Rivoirard, and Gilles Stoltz) taught the following classes.

Licence : Statistiques, 39h, niveau L2, Université Paris-Dauphine, par Vincent Rivoirard

Licence : Apprentissage, 20h, niveau L3, Ecole normale supérieure, par Olivier Catoni et Gilles Stoltz

Licence : Théorie des probabilités, 40h, niveau L3, ISUP – Université Pierre et Marie Curie), par Gérard Biau

Licence : Statistiques pour citoyens d'aujourd'hui et managers de demain, 40h, niveau L3, HEC Paris, par Gilles Stoltz

Master : Groupe de travail en statistique, 12h, niveau M1, Ecole normale supérieure, par Gérard Biau, Olivier Catoni et Gilles Stoltz

Master : Statistique mathématique, 30h, niveau M1, Ecole normale supérieure, par Gérard Biau

Master : Statistique non-paramétrique, 8h, niveau M1, Ecole normale supérieure, par Vincent Rivoirard

Master : Statistique non-paramétrique, 35h, niveau M1, Université Paris-Dauphine, par Vincent Rivoirard

Master : Classification et statistique en grandes dimensions, 18h, niveau M2, Université Paris-Sud, par Vincent Rivoirard

Master : Statistiques et théorie de l'information, 10h, niveau M2, Université Paris-Sud, par Gilles Stoltz

Master : Apprentissage statistique, 36h, niveau M2, Université Pierre et Marie Curie, par Gérard Biau

Master : Méthodes pour les modèles de régression, 21h, niveau M2, Université Paris-Dauphine, par Vincent Rivoirard

Master : Statistique bayésienne non-paramétrique, 21h, niveau M2, Université Paris-Dauphine, par Vincent Rivoirard

Master : Examinateur à l'oral de probabilités et statistiques de l'agrégation de mathématiques, par Gilles Stoltz

PhD & HdR

HdR : Gilles Stoltz, *Contributions à la prévision séquentielle de suites arbitraires : applications à la théorie des jeux répétés et études empiriques des performances de l'agrégation d'experts*, Université Paris-Sud; defended on February 3, 2011

PhD : Sébastien Gerchinovitz, *Prédiction de suites individuelles et cadre statistique classique : étude de quelques liens autour de la régression parcimonieuse et des techniques d'agrégation*, Université Paris-Sud; defended on December 12, 2011; supervised by Gilles Stoltz

PhD in progress : Thomas Mainguy, *Modèles statistiques pour la linguistique computationnelle*, since September 2009, supervised by Olivier Catoni

PhD in progress : Pierre Gaillard, since September 2011, supervised by Gilles Stoltz

PhD in progress : Emilien Joly, since September 2011, supervised by Gábor Lugosi and co-supervised by Gilles Stoltz

Several other PhD in progress : Gérard Biau and Vincent Rivoirard [co-]supervise[d] several other PhD students who are not members of our project-team (respectively, Benjamin Auder, Aurélie Fischer, Benoît Patra, Clément Levrard, Benjamin Guedj, Svetlana Gribkova, Baptiste Gregorutti for Gérard Biau, and Laure Sansonnet for Vincent Rivoirard)

MSc thesis: Pierre Gaillard (Master MVA, ENS Cachan) was supervised by Gilles Stoltz during this MSc thesis, whose subject was the use of aggregation techniques (based on random forests and/or steming from the theory of the prediction of individual sequences) for the forecasting of electricity consumption.

9. Bibliography

Major publications by the team in recent years

[1] J.-Y. Audibert, O. Catoni. *Robust linear least squares regression*, in "The Annals of Statistics", 2011, in press, http://hal.inria.fr/hal-00522534.

[2] K. Bertin, E. Le Pennec, V. Rivoirard. *Adaptive Dantzig density estimation*, in "Annales de l'IHP, Probabilités et Statistiques", 2011, vol. 47, n° 1, p. 43–74, http://hal.inria.fr/hal-00381984/en.

[3] G. Biau, L. Devroye, G. Lugosi. *Consistency of random forests and other averaging classifiers*, in "Journal of Machine Learning Research", 2008, vol. 9, p. 2015–2033.

[4] G. Biau, L. Devroye, G. Lugosi. *On the performance of clustering in Hilbert spaces*, in "IEEE Transactions on Information Theory", 2008, vol. 54, p. 781–790.

[5] O. Catoni. *Statistical Learning Theory and Stochastic Optimization — Lectures on Probability Theory and Statistics, École d'Été de Probabilités de Saint-Flour XXXI – 2001*, Lecture Notes in Mathematics, Springer, 2004, vol. 1851, 269 pages.

[6] O. Catoni. *PAC-Bayesian Supervised Classification: The Thermodynamics of Statistical Learning*, IMS Lecture Notes Monograph Series, Institute of Mathematical Statistics, 2007, vol. 56, 163 pages, http://dx.doi.org/10.1214/074921707000000391.

[7] G. Lugosi, S. Mannor, G. Stoltz. *Strategies for prediction under imperfect monitoring*, in "Mathematics of Operations Research", 2008, vol. 33, p. 513–528.

[8] B. Mauricette, V. Mallet, G. Stoltz. *Ozone ensemble forecast with machine learning algorithms*, in "Journal of Geophysical Research", 2009, vol. 114, D05307, http://dx.doi.org/10.1029/2008JD009978.

[9] V. Rivoirard, G. Stoltz. *Statistique en action*, Vuibert, 2009, Main volume contains 320 pages, supplementary material of 220 pages is available for free download.

Publications of the year

Doctoral Dissertations and Habilitation Theses

[10] S. GERCHINOVITZ. *Prédiction de suites individuelles et cadre statistique classique : étude de quelques liens autour de la régression parcimonieuse et des techniques d'agrégation*, Université Paris-Sud, December 2011, http://www.math.ens.fr/~gerchinovitz/docs/These-Gerchinovitz.pdf.

[11] G. STOLTZ. *Contributions à la prévision séquentielle de suites arbitraires : applications à la théorie des jeux répétés et études empiriques des performances de l'agrégation d'experts*, Université Paris-Sud, February 2011, Habilitation à diriger des recherches, http://www.math.ens.fr/~stoltz/HDR/HDR-Stoltz-Eng-Web.pdf.

Articles in International Peer-Reviewed Journal

[12] J.-Y. AUDIBERT, O. CATONI. *Robust linear least squares regression*, in "The Annals of Statistics", 2011, in press, http://hal.inria.fr/hal-00522534.

[13] K. BERTIN, E. LE PENNEC, V. RIVOIRARD. *Adaptive Dantzig density estimation*, in "Annales de l'IHP, Probabilités et Statistiques", 2011, vol. 47, n⁰ 1, p. 43–74, http://hal.inria.fr/hal-00381984/en.

[14] G. BIAU. *Analysis of a random forests model*, in "Journal of Machine Learning Research", 2011, in press, http://hal.inria.fr/hal-00476545/en.

[15] G. BIAU, F. CHAZAL, D. COHEN-STEINER, L. DEVROYE, C. RODRÍGUEZ. *A weighted k–nearest neighbor density estimate for geometric inference*, in "Electronic Journal of Statistics", 2011, vol. 5, p. 204–237, http://hal.inria.fr/inria-00560623/en.

[16] G. BIAU, A. FISCHER. *Parameter selection for principal curves*, in "IEEE Transactions on Information Theory", 2011, in press, http://hal.inria.fr/hal-00565540/en.

[17] G. BIAU, A. MAS. *PCA–kernel estimation*, in "Statistics & Risk Modeling", 2011, in press, http://hal.inria.fr/hal-00467013/en.

[18] S. BUBECK, R. MUNOS, G. STOLTZ. *Pure Exploration in Finitely-Armed and Continuous-Armed Bandit Problems*, in "Theoretical Computer Science", 2011, vol. 412, p. 1832–1852, http://hal.inria.fr/hal-00257454/en.

[19] S. BUBECK, R. MUNOS, G. STOLTZ, C. SZEPESVÁRI. *X-Armed Bandits*, in "Journal of Machine Learning Research", April 2011, vol. 12, p. 1655–1695, http://hal.inria.fr/hal-00450235/en.

[20] O. CATONI. *Challenging the empirical mean and empirical variance: a deviation study*, in "Annales de l'IHP, Probabilités et Statistiques", 2011, in press; with an online supplement, http://hal.inria.fr/hal-00517206/en.

[21] M. DOUMIC, M. HOFFMANN, P. REYNAUD-BOURET, V. RIVOIRARD. *Nonparametric estimation of the division rate of a size-structured population*, in "SIAM Journal on Numerical Analysis", 2011, in press, http://hal.inria.fr/hal-00578694/en.

[22] B. PATRA, G. BIAU. *Sequential quantile prediction of time series*, in "IEEE Transactions on Information Theory", 2011, vol. 57, p. 1664–1674, http://hal.inria.fr/hal-00410120/en.

[23] G. STOLTZ, T. MICHALSKI. *Do countries falsify economic data strategically? Some evidence that they might.*, in "The Review of Economics and Statistics", 2011, in press, http://hal.inria.fr/halshs-00482106/en.

International Conferences with Proceedings

[24] S. BUBECK, G. STOLTZ, J. Y. YU. *Lipschitz Bandits without the Lipschitz Constant*, in "ALT 2011 – 22nd International Conference on Algorithmic Learning Theory", Espoo, Finland, J. KIVINEN, C. SZEPESVÁRI, E. UKKONEN, T. ZEUGMANN (editors), LNCS / Lecture Notes in Artificial Intelligence, Springer, October 2011, http://hal.inria.fr/hal-00595692/en.

[25] S. GERCHINOVITZ. *Sparsity regret bounds for individual sequences in online linear regression*, in "COLT 2011 – The 24rd Annual Conference on Learning Theory", Budapest, Hungary, S. KAKADE, U. VON LUXBURG (editors), OmniPress, July 2011, http://hal.inria.fr/inria-00552267/en.

[26] S. GERCHINOVITZ, J. Y. YU. *Adaptive and Optimal Online Linear Regression on L_1–balls*, in "ALT 2011 – 22nd International Conference on Algorithmic Learning Theory", Espoo, Finland, J. KIVINEN, C. SZEPESVÁRI, E. UKKONEN, T. ZEUGMANN (editors), LNCS / Lecture Notes in Artificial Intelligence, Springer, October 2011, http://hal.inria.fr/hal-00594399/en.

[27] O.-A. MAILLARD, R. MUNOS, G. STOLTZ. *A Finite-Time Analysis of Multi-armed Bandits Problems with Kullback-Leibler Divergences*, in "COLT 2011 – The 24rd Annual Conference on Learning Theory", Budapest, Hungary, S. KAKADE, U. VON LUXBURG (editors), OmniPress, July 2011, http://hal.inria.fr/inria-00574987/en.

[28] S. MANNOR, V. PERCHET, G. STOLTZ. *Robust approachability and regret minimization in games with partial monitoring*, in "COLT 2011 – The 24rd Annual Conference on Learning Theory", Budapest, Hungary, S. KAKADE, U. VON LUXBURG (editors), OmniPress, July 2011, http://hal.inria.fr/hal-00595695/en.

Books or Proceedings Editing

[29] P. ALQUIER, E. GAUTIER, G. STOLTZ (editors). *Inverse Problems and High-Dimensional Estimation*, Lecture Notes in Statistics, Springer, 2011, vol. 203, Stats in the Château summer school (revised lectures), http://www.springer.com/statistics/book/978-3-642-19988-2.

Other Publications

[30] P. ALQUIER, G. BIAU. *Sparse single-index model*, 2011, under review by Journal of Machine Learning Research, http://hal.inria.fr/hal-00556652/en.

[31] G. BIAU, Y. YATRACOS. *On the shrinkage estimation of variance and Pitman closeness criterion*, 2011, submitted to Journal de la Société Francaise de Statistiques, http://www.lsta.upmc.fr/BIAU/by.pdf.

[32] M. DEVAINE, Y. GOUDE, G. STOLTZ. *Forecasting the electricity consumption by aggregating specialized experts; a review of the sequential aggregation of specialized experts, with an application to Slovakian and French country-wide one-day-ahead (half-)hourly predictions*, 2011, under review by Machine Learning, http://hal.inria.fr/hal-00484940/en.