



IN PARTNERSHIP WITH:
CNRS

**Ecole normale supérieure de
Paris**

Activity Report 2013

Project-Team CLASSIC

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

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

RESEARCH CENTER
Paris - Rocquencourt

THEME
**Optimization, machine learning and
statistical methods**

Table of contents

1. Members	1
2. Overall Objectives	1
3. Research Program	2
3.1. Regression models of supervised learning	2
3.1.1. PAC-Bayes inequalities	2
3.1.2. Sparsity and ℓ_1 -regularization	2
3.1.3. Pushing it to the extreme: no assumption on the data	2
3.2. On-line aggregation of predictors for the prediction of time series, with or without stationarity assumptions	2
3.3. Multi-armed bandit problems, prediction with limited feedback	3
3.3.1. Bandit problems	3
3.3.2. A generalization of the regret: the approachability of sets	3
4. Application Domains	3
4.1. Forecasting of the electricity consumption	3
4.2. Forecasting of the air quality	3
4.3. Forecasting of the production data of oil reservoirs	3
4.4. Forecasting of exchange rates	4
4.5. Data mining, massive data sets	4
4.6. Computational linguistics	4
4.7. Statistical inference on biological data	4
5. New Results	4
5.1. Contributions earlier to 2013 but only published in 2013	4
5.2. Approachability with partial monitoring	4
5.3. High-dimensional learning and complex data	4
5.4. Dimension free principal component analysis	5
5.5. Statistical models for corpus linguistics	5
6. Bilateral Contracts and Grants with Industry	6
7. Partnerships and Cooperations	6
7.1. National Initiatives	6
7.2. International Initiatives	6
8. Dissemination	6
8.1. Scientific Animation	6
8.1.1. Conference organization	6
8.1.2. Organization of seminars	7
8.1.3. Editorial activities, reports written on articles	7
8.1.4. Participation to national or local evaluation or recruitment committees, to scientific societies	7
8.1.5. Honors and distinctions	7
8.2. Teaching - Supervision - Juries	7
8.2.1. Teaching	7
8.2.2. Supervision	8
8.2.3. Juries	8
8.3. Popularization	8
9. Bibliography	8

Project-Team CLASSIC

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

The team is located at Ecole normale supérieure, 45 rue d'Ulm, Paris.

Creation of the Team: 2009 July 01, updated into Project-Team: 2010 January 01.

1. Members

Research Scientists

Olivier Catoni [Team leader, CNRS, Senior Researcher, HdR]
Gilles Stoltz [CNRS, Researcher, until Aug 2013, HdR]

Faculty Member

G erard Biau [Univ. Paris Pierre-et-Marie-Curie, Professor, HdR]

External Collaborator

Vincent Rivoirard [Univ. Paris-Dauphine, HdR]

PhD Students

Paul Baudin [Inria, until Aug 2013]
Pierre Gaillard [EDF, until Aug 2013]
Ilaria Giulini [ENS Paris]
Emilien Joly [Univ. Paris XI]
Thomas Mainguy [ENS Paris]

Administrative Assistant

Lindsay Polienor [Inria]

Other

Christophe Amat [Inria, M.Sc. intern, from Apr 2013 until Jul 2013]

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);
- statistical modeling applied to linguistics, statistical inference of grammar models.

We are involved in the following applications:

- the improvement of prediction through the on-line aggregation of predictors, with an emphasis on the forecasting of air quality, electricity consumption, production data of oil reservoirs, exchange rates;
- natural image analysis, and more precisely the use of unsupervised learning in data representation;
- computational linguistics;
- statistical inference on biological and neurobiological data.

3. Research Program

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) , $i = 1, \dots, n$, 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] and on principal component analysis of a random sample of points in high dimension. It has introduced the idea of combining the PAC-Bayesian approach with the use of thresholded exponential moments [7], 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 certain cases, we may formulate absolutely no assumption on the data (x_i, y_i) , $i = 1, \dots, n$, 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 goal here is to upper bound the regret by a quantity as small as possible.

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

We are interested 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. Forecasting of the production data of oil reservoirs

Our partner is IFP Energies nouvelles. The goal is to aggregate in a sequential fashion the forecasts made by some (about 100) base experts in order to predict some behaviors (gas/oil ratio, cumulative oil extracted, water cut) of the exploitation of some oil wells.

4.4. Forecasting of exchange rates

Our partner is HEC Paris. The goal is to aggregate in a sequential fashion the forecasts made by some (about 5) base macro-economic variables to predict monthly-averaged exchange rates.

4.5. Data mining, massive data sets

Our partner is the start-up Safety Line. The purpose of this application is to investigate statistical learning strategies for mining massive data sets originated from aircraft high-frequency recordings and improve security.

4.6. Computational linguistics

We propose and study new language models that bridge the gap between models oriented towards the statistical analysis of large corpora and grammars oriented towards the description of syntactic features as understood by academic experts. We have conceived a new kind of grammar, based on some cut and paste mechanism and some label aggregation principle, that can be fully learnt from a corpus. We are currently testing this model and studying its mathematical properties and relations with some other new statistical models based on conditional independence assumptions.

4.7. Statistical inference on biological data

The question is about understanding how interactions between neurons can be detected. A mathematical modeling is given by multivariate Hawkes processes. Lasso-type methods can then be used to estimate interaction functions in the nonparametric setting by using fast algorithms, providing inference of the unitary event activity of individual neurons.

5. New Results

5.1. Contributions earlier to 2013 but only published in 2013

Participants: Gérard Biau, Pierre Gaillard, Gilles Stoltz.

We do not discuss here the contributions provided by [14], [12], [13], [16] since they were achieved in 2012 or earlier (but only published this year due to the reviewing and publishing process).

5.2. Approachability with partial monitoring

Participant: Gilles Stoltz.

This line of research has been developed in our team since its creation (see, in particular, the founding article [9] as well as several other publications in the previous reports). Following the earlier contribution on exhibiting an efficient algorithm for approachability with partial monitoring based on some necessary and sufficient dual condition, we study in [15] the primal approach: the statement of the condition and the existence of (efficient or inefficient) algorithms based on it.

5.3. High-dimensional learning and complex data

Participant: Gérard Biau.

We describe four (not so related) contributions on the theme of high-dimensional learning and complex data. In [17] we address the problem of supervised classification of Cox process trajectories, whose random intensity is driven by some exogenous random covariable. The classification task is achieved through a regularized convex empirical risk minimization procedure, and a nonasymptotic oracle inequality is derived. The results are obtained by taking advantage of martingale and stochastic calculus arguments, which are natural in this context and fully exploit the functional nature of the problem.

The cellular tree classifier model addresses a fundamental problem in the design of classifiers for a parallel or distributed computing world: Given a data set, is it sufficient to apply a majority rule for classification, or shall one split the data into two or more parts and send each part to a potentially different computer (or cell) for further processing? At first sight, it seems impossible to define with this paradigm a consistent classifier as no cell knows the “original data size”, n . However, we show in [18] that this is not so by exhibiting two different consistent classifiers.

A new method for combining several initial estimators of the regression function is introduced. Instead of building a linear or convex optimized combination over a collection of basic estimators r_1, \dots, r_M , [19] uses them as a collective indicator of the proximity between the training data and a test observation. This local distance approach is model-free and very fast. More specifically, the resulting collective estimator is shown to perform asymptotically at least as well in the L^2 sense as the best basic estimator in the collective. A companion R package called COBRA (standing for COMBined Regression Alternative) is presented (downloadable on <http://cran.r-project.org/web/packages/COBRA/index.html>). Substantial numerical evidence is provided on both synthetic and real data sets to assess the excellent performance and velocity of the method in a large variety of prediction problems.

The impact of letting the dimension d go to infinity on the L^p -norm of a random vector with i.i.d. components has surprising consequences, which may dramatically affect high-dimensional data processing. This effect is usually referred to as the *distance concentration phenomenon* in the computational learning literature. Despite a growing interest in this important question, previous work has essentially characterized the problem in terms of numerical experiments and incomplete mathematical statements. In the paper [20], we solidify some of the arguments which previously appeared in the literature and offer new insights into the phenomenon.

5.4. Dimension free principal component analysis

Participants: Olivier Catoni, Ilaria Giulini.

In a work in progress, Ilaria Giulini, as part of her PhD studies, proved the following dimension free inequality, related to Principal Component Analysis in high dimension. Given an i.i.d. sample X_i , $1 \leq i \leq n$ of vector valued random variables $X_i \in \mathbf{R}^d$, there exists an estimator \widehat{N} of the quadratic form $N(\theta) = \mathbf{E}(\langle \theta, X \rangle^2)$ such that for any $n \leq 10^{20}$, with probability at least $1 - 2\epsilon$, for any $\theta \in \mathbf{R}^d$,

$$\mathbf{1}(4\mu < 1) \left| \frac{\widehat{N}(\theta)}{N(\theta)} - 1 \right| \leq \frac{\mu}{1 - 4\mu},$$

where

$$\mu = \sqrt{\frac{2.07(\kappa - 1)}{n} \left[\log(\epsilon^{-1}) + 4.3 + \frac{1.6 \|\theta\|^2 \mathbf{Tr}(G)}{N(\theta)} \right]} + \sqrt{\frac{184 \kappa \|\theta\|^2 \mathbf{Tr}(G)}{nN(\theta)}},$$

where $G = \mathbf{E}(XX^\top)$ is the Gram matrix and where $\kappa = \sup \left\{ \frac{\mathbf{E}(\langle \theta, X \rangle^4)}{\mathbf{E}(\langle \theta, X \rangle^2)^2}, \theta \in \mathbf{R}^d \setminus \mathbf{Ker}(G) \right\}$ is some

kurtosis coefficient. This result proves that the expected energy in direction θ can be estimated at a rate that is independent of the dimension of the ambient space \mathbf{R}^d . It is obtained using PAC-Bayes inequalities with Gaussian parameter perturbations. The same bound holds in a Hilbert space of infinite dimension, opening the possibility of a rigorous mathematical study of kernel principal component analysis of random data, where the data are represented in a possibly infinite dimensional reproducing kernel Hilbert space.

5.5. Statistical models for corpus linguistics

Participants: Olivier Catoni, Thomas Mainguy.

In [21] we describe a language model as the invariant measure of a Markov chain on sentence samples. The kernel of this Markov chain is defined with the help of some context free grammars : from the sentence sample, a random parse model produces a context free grammar with weighted rules, and from this grammar, a new sentence sample is formed by applying the rules randomly. We prove various mathematical properties of this Markov process, related to its computation cost and the fact that it is weakly reversible and therefore ergodic on each of its communicating classes. As a companion to the Markov chain on sentence samples, we can also define a Markov chain on weighted context free grammars. This leads to another type of grammar, that we called Toric Grammars, defined by a family of context tree grammars that can be computed from any of its members as the communicating class of a Markov chain on context free grammars with weighted rules. Preliminary simulations on small data sets are very encouraging, in that they show that this type of model is able to grasp the recursive nature of natural languages.

6. Bilateral Contracts and Grants with Industry

6.1. Bilateral Contracts with Industry

An industrial contract with EDF R&D (cf. CIFRE PhD of Pierre Gaillard) has come into effect as of November 8, 2012, and will last 3 years.

7. Partnerships and Cooperations

7.1. National Initiatives

- ANR project in the blank program: Calibration (2012–2015; involves Vincent Rivoirard, who is the coordinator; see <https://sites.google.com/site/anrcalibration/home>)
- ANR project in the blank program: Banhdits (2010–2013; involves Vincent Rivoirard; see <https://sites.google.com/site/anrcalibration/home>)
- PEPS Bio-Maths (“Estimation de graphes de dépendance entre neurones thalamiques et cortico-thalamiques via des modèles de Hawkes multivariés; 2012–2013; involves Vincent Rivoirard)

7.2. International Initiatives

We have one formal international collaboration, with

- Karine Bertin, University of Valparaiso, Chile (International cooperation CONICYT project, Andes Foundation project);

and other informal ones:

- Luc Devroye, McGill University, Canada;
- David Mason, Delaware University, USA;
- Shie Mannor, Technion, Israel.

8. Dissemination

8.1. Scientific Animation

8.1.1. Conference organization

Gérard Biau co-organized a one-day workshop on “High-dimensional clustering” at Ecole normale supérieure (September 2013).

8.1.2. Organization of seminars

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>).

8.1.3. Editorial activities, reports written on articles

Gérard Biau is (co-)Editor-in-Chief of *ESAIM: Probability and Statistics* and serves as an Associate Editor for the journals *Annales de l'ISUP* and *International Statistical Review*.

Vincent Rivoirard serves as an Associate Editor for *ESAIM: Probability and Statistics*.

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

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

Vincent Rivoirard is the General Secretary of SFdS and Gérard Biau is the Vice-President of this society.

Gérard Biau is a member of the national board of French universities (CNU) within the applied mathematics section (number 26).

Olivier Catoni is a member of the doctoral commission in mathematics of University Pierre et Marie Curie.

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

8.1.5. Honors and distinctions

Gérard Biau is a member of the Institut Universitaire de France (IUF).

8.2. Teaching - Supervision - Juries

8.2.1. Teaching

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

Licence : Olivier Catoni, Apprentissage, 10h, niveau L3, Ecole normale supérieure, France

Licence : Gilles Stoltz, Statistiques pour citoyens d'aujourd'hui et managers de demain, 60h, niveau L3, HEC Paris, France

Master : Gérard Biau and Olivier Catoni, Groupe de travail Statistique, 20h, niveau M1, Ecole normale supérieure, France

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

Master : Vincent Rivoirard, Leçons de mathématiques, niveau M1, 10h, Ecole normale supérieure, France

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

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

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

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

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

Master : Gilles Stoltz, Examineur à l'oral de probabilités et statistiques de l'agrégation de mathématiques, France

8.2.2. Supervision

We only indicate the on-going PhD theses of members of our team.

PhD in progress : Thomas Mainguy, Statistical learning in computational linguistics, since September 2010, supervised by Olivier Catoni

PhD in progress : Emilien Joly, Phase transition of optimal risk and detection of contamination, since September 2011, supervised by Gábor Lugosi and co-supervised by Gilles Stoltz

PhD in progress : Pierre Gaillard, Aggregation of specialized predictors for the forecasting of electricity consumption, since September 2011, supervised by Gilles Stoltz

PhD in progress : Ilaria Giulini, Dimension free PAC-Bayes bounds for the Gram matrix and unsupervised clustering on the sphere of a Reproducing Kernel Hilbert space, since September 2012, supervised by Olivier Catoni

PhD in progress : Paul Baudin, Robust aggregation of predictors for the forecasting of air quality, with measures of uncertainties, since October 2012, supervised by Gilles Stoltz and co-supervised by Vivien Mallet

8.2.3. Juries

Gérard Biau, Olivier Catoni and Vincent Rivoirard participated in different PhD/HdR defense juries, sometimes as external reviewers.

8.3. Popularization

Gilles Stoltz

- wrote a blog post for MPT2013 and an article for The Huffigton Post, in the researchers' blog;
- gave a conference for high school teachers in Orléans;
- gave a mini-course on statistics to "classes préparatoires" professors at ENS Paris and a mini-course of game theory at ENS Rennes, targeted to inspectors of high-school-teachers.

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, vol. 39, n^o 5, pp. 2766-2794, <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^o 1, pp. 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, pp. 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, pp. 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 p. , <http://dx.doi.org/10.1214/074921707000000391>
- [7] O. CATONI. *Challenging the empirical mean and empirical variance: A deviation study*, in "Annales de l'Institut Henri Poincaré - Probabilités et Statistiques", 2012, vol. 48, n^o 4, pp. 1148-1185
- [8] M. DEVAINE, P. GAILLARD, Y. GOUDE, G. STOLTZ. *Forecasting 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*, in "Machine Learning", 2012, to appear
- [9] G. LUGOSI, S. MANNOR, G. STOLTZ. *Strategies for prediction under imperfect monitoring*, in "Mathematics of Operations Research", 2008, vol. 33, pp. 513–528
- [10] B. MAURICETTE, V. MALLET, G. STOLTZ. *Ozone ensemble forecast with machine learning algorithms*, in "Journal of Geophysical Research", 2009, vol. 114, <http://dx.doi.org/10.1029/2008JD009978>
- [11] V. RIVOIRARD, G. STOLTZ. , *Statistique mathématique en action*, second edition, Vuibert, 2012, <http://www.dma.ens.fr/statenaction/>

Publications of the year

Articles in International Peer-Reviewed Journals

- [12] P. ALQUIER, G. BIAU. *Sparse single-index model*, in "Journal of Machine Learning Research", 2013, vol. 14, pp. 243–280, <http://hal.inria.fr/hal-00556652>
- [13] O. CAPPÉ, A. GARIVIER, O.-A. MAILLARD, R. MUNOS, G. STOLTZ. *Kullback-Leibler Upper Confidence Bounds for Optimal Sequential Allocation*, in "Annals of Statistics", 2013, vol. 41, n^o 3, pp. 1516-1541, Accepted, <http://hal.inria.fr/hal-00738209>
- [14] M. DEVAINE, P. GAILLARD, Y. GOUDE, G. STOLTZ. *Forecasting electricity consumption by aggregating specialized experts*, in "Machine Learning", 2013, vol. 90, n^o 2, pp. 231-260, <http://hal.inria.fr/hal-00484940>
- [15] S. MANNOR, V. PERCHET, G. STOLTZ. *A Primal Condition for Approachability with Partial Monitoring*, in "Journal of Dynamics and Games", 2013, in press, <http://hal.inria.fr/hal-00772056>
- [16] T. MICHALSKI, G. STOLTZ. *Do countries falsify economic data strategically? Some evidence that they might.*, in "The Review of Economics and Statistics", May 2013, vol. 95, n^o 2, pp. 591-616 [DOI : 10.1162/REST_A_00274], <http://hal.inria.fr/halshs-00482106>

Other Publications

- [17] G. BIAU, B. CADRE, Q. PARIS. , *Cox Process Learning*, 2013, <http://hal.inria.fr/hal-00820838>
- [18] G. BIAU, L. DEVROYE. , *Cellular Tree Classifiers*, 2013, <http://hal.inria.fr/hal-00778520>

- [19] G. BIAU, A. FISCHER, B. GUEDJ, J. MALLEY. , *COBRA: A Nonlinear Aggregation Strategy*, 2013, 40 p. , 5 tables, 12 figures, <http://hal.inria.fr/hal-00798579>
- [20] G. BIAU, D. MASON. , *High-dimensional p-norms*, 2013, 19 p. , <http://hal.inria.fr/hal-00879436>
- [21] O. CATONI, T. MAINGUY. , *Toric grammars: a new statistical approach to natural language modeling*, 2013, <http://hal.inria.fr/hal-00934487>