



INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

Project-Team classic

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

Paris - Rocquencourt

Theme : Optimization, Learning and Statistical Methods

Activity
R *eport*

2010

Table of contents

1. Team	1
2. Overall Objectives	1
3. Scientific Foundations	2
3.1. Regression models of supervised learning	2
3.2. On-line aggregation of predictors for the prediction of time series, with or without stationarity assumptions	2
3.3. Sparsity and ℓ_1 -regularization	2
3.4. Multi-armed bandit problems	3
3.5. Applications to the theory of repeated games	3
4. Application Domains	3
4.1. Forecasting of the electricity consumption	3
4.2. Forecasting of the air quality	4
4.3. Management of the supply chain	4
4.4. Computational linguistics	4
4.5. Statistical inference on genomic data	4
5. New Results	4
5.1. Contributions earlier to 2010 but only published in 2010	4
5.2. Supervised statistical inference: regression and classification	4
5.3. Sparse regression estimation	5
5.4. Advances in the application of individual sequences to operational forecasting	5
5.5. On-line aggregation of predictors for the prediction of time series	6
5.6. Calibrated forecasts	6
6. Contracts and Grants with Industry	6
7. Other Grants and Activities	6
7.1. National Initiatives	6
7.2. European Initiatives	7
7.3. International Initiatives	7
8. Dissemination	7
8.1. Doctoral studies	7
8.2. Invited conferences	7
8.3. Editorial responsibilities	8
8.4. Animation of the scientific community	8
8.5. Teaching	8
9. Bibliography	8

This project is a common project with CNRS and Ecole normale supérieure. The team has been created on July the 1st, 2009 and became an INRIA project on January the 1st, 2010.

1. Team

Research Scientists

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

Faculty Members

G rard Biau [Professor, Universit  Paris Pierre-et-Marie-Curie, HdR]
Vincent Rivoirard [Professor, Universit  Paris-Dauphine, HdR]

PhD Students

S bastien Gerchinovitz [PhD student, fellow of Universit  Paris-Sud]
Thomas Mainguy [PhD student, student at Ecole normale sup rieure]

Post-Doctoral Fellow

Jia Yuan Yu [Post-doc student, funded by an ANR grant]

Administrative Assistant

H l ne Milome [Assistant; 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 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 genomic data by using sparse statistical regression methods.

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 $(X_i, Y_i), i = 1, \dots, n$, 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.

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 [5], then on regression (see the papers [18], [19] discussed below). 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.

Another line of research in regression estimation is the use of sparse models, and its link with ℓ_1 -regularization. Selecting a few variables from a large set of candidates in a computationally efficient way is a major challenge of statistical learning. Another approach to catch more general situations, is to predict outputs in a sequential way. If a cumulated loss is considered, this can be done even under weaker assumptions than what is possible within the regression framework. These two lines are described in the next two items.

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

Here, we are concerned 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. Sparsity and ℓ_1 -regularization

Extensively used approaches in modern nonparametric statistics for the problems of estimation, prediction, or model selection, are based on *regularization*. The joint minimization of some empirical criterion and some penalty function 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.

The Lasso algorithm, however, needs a tuning parameter, which is to be calibrated. However, the parameters that are good in theory (the ones that are used to derive sharp oracle inequalities) are in general too conservative for practical purposes. Our primary aim is to exhibit a calibration procedure for stochastic data that ensures both good practical and theoretical performance.

A secondary aim is to have a theoretical analysis of the Lasso in the context of individual sequences.

3.4. Multi-armed bandit problems

This is a stochastic problem, in which 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.

Typical results in the literature are 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 take 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 is to have an adaptive procedure, that, given one unknown environment (with unknown regularity), ensures that the regret is asymptotically small; it would be even better to control the regret in some uniform manner (in a distribution-free sense up to the regularity parameters).

3.5. Applications to the theory of repeated games

The individual sequences mentioned above can be thought of as being chosen by an opponent player, in which case a repeated two-player game is at hand. We study two fundamental tools in this context: calibration and approachability.

Calibration is the ability to forecast well, on average, the opponent's actions. It is often used as the property of some auxiliary strategy on which some main strategy can be built. The latter will turn to be efficient since it can accurately reconstruct the opponent's behavior.

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. We want to extend the result 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 which could hopefully 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. Combining ideas from variable-order Markov chains and lossless compression schemes of the Lempel-Ziv family, a new model is presently under construction, which should derive syntactic patterns using as few observations as possible. (Note: this application was not present in the project we submitted to create the team; it is dealt with by Thomas Mainguy, who started on September 2010 a thesis about corpus linguistics, supervised by Olivier Catoni.)

4.5. Statistical inference on genomic data

Human genome is composed of about 30 000 genes, which may be transcribed in about 160 000 different expressions; to understand how transcription 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 should help overcoming this numerical issue.

5. New Results

5.1. Contributions earlier to 2010 but only published in 2010

Participants: Gérard Biau, Gilles Stoltz.

We do not discuss here the contributions provided by [15], [10], [9], [12], [11], since they were achieved in 2009 or earlier (but only published this year due to long queues in publication tracks of journals).

5.2. Supervised statistical inference: regression and classification

Participants: Gérard Biau, Olivier Catoni.

Least square regression with random design is a central issue in supervised statistical inference. The team, in collaboration with Willow, was able in [18] to show on the one hand that the ordinary least square estimator has an asymptotic rate optimal behaviour proportional to d/n , where d is the dimension and n the sample size, under very weak assumptions (existence of a quadratic moment for the noise and of a fourth moment for the design, without any assumptions on the conditioning of the Gram matrix). Moreover, this result can be extended to ridge regression, the dimension being replaced with some lower *effective ridge dimension*. However, under such hypotheses, this asymptotic regime can be reached arbitrarily slowly. To obtain non asymptotic bounds, it is necessary to make the estimator itself more robust. This is possible through some min-max truncation scheme, for which it is possible to give a non asymptotic convergence rate depending only on the kurtosis of a few quantities. This min-max scheme is feasible in practice, involving in experiments a load of computations of order 50 times what is needed for the ordinary least square estimator. Experiments also show improved performance in comparison with the ordinary least square estimator, when the noise is heavy tailed, and preserved performances otherwise (where the two estimators compute de same solution).

In order to use PAC-Bayes inequalities, it is necessary to consider a perturbation of the parameter, in the form of a posterior distribution. For this reason, the theory gives sharper results [19] for randomized and quite involved estimators, defined by posterior distributions. Using this kind of estimators, it is possible to show non asymptotic range optimal rates for general loss functions under even milder dimension and margin assumptions (generalizing the notion of margin introduced by Mammen and Tsybakov).

On the other hand, the min-max truncation scheme proposed for least square estimation can be simplified in the case of mean estimation [22], leading to a mean estimator with better deviation properties than the empirical mean estimator for heavy tailed distributions (such as the mixture of two Gaussian measures with different standard deviations).

Another direction of research to turn statistical regression into a learning tool is to find efficient ways to deal with high dimension inputs. Various aggregation and dimension reductions methods have been studied within the team—among which random forests, which we discuss below, and PCA-Kernel estimation, which we discuss now. Indeed, many statistical estimation techniques for high-dimensional or functional data are based on a preliminary dimension reduction step, which consists in projecting the sample $\mathbf{X}_1, \dots, \mathbf{X}_n$ onto the first D eigenvectors of the Principal Component Analysis (PCA) associated with the empirical projector $\hat{\Pi}_D$. Classical nonparametric inference methods such as kernel density estimation or kernel regression analysis are then performed in the (usually small) D -dimensional space. However, the mathematical analysis of this data-driven dimension reduction scheme raises technical problems, due to the fact that the random variables of the projected sample $(\hat{\Pi}_D \mathbf{X}_1, \dots, \hat{\Pi}_D \mathbf{X}_n)$ are no more independent. As a reference for further studies, we offer in the paper [21] several results showing the asymptotic equivalencies between important kernel-related quantities based on the empirical projector and its theoretical counterpart.

5.3. Sparse regression estimation

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

The paper [24] by Sébastien Gerchinovitz imports the notion of sparse oracle inequalities within the theory of individual sequences. It is based on a forecaster –SEW– that Dalalyan and Tsybakov introduced in a series of articles, the first of them being presented at COLT’07, and studied in a stochastic (i.i.d.) setting. The forecaster relies on some tuning parameters and the question of their adaptive calibration with respect in particular to the variance was left open. However, the individual sequence bounds proved on its extension naturally imply stochastic bounds; and since the individual sequence version of SEW is perfectly calibrated on-line, it solves the question left open therein. The mathematical techniques used to prove the extension are in particular a PAC-Bayesian inequality developed by Olivier Catoni and an adaptive exponentially weighted average scheme exhibited by Gilles Stoltz and co-authors of his. Sébastien Gerchinovitz also benefited from the background and some advice of Vincent Rivoirard, with respect to the stochastic scenario.

Another line of research in this context was performed by Gérard Biau in [20], [13] and is concerned with random forests. These are a scheme proposed by Leo Breiman for building a predictor ensemble with a set of decision trees that grow in randomly selected subspaces of data. Despite growing interest and practical use, there had been little exploration of the statistical properties of random forests, and little is known about the mathematical forces driving the algorithm. In this respect he shows in particular that a variant (proposed by Breiman and his co-authors) of the base procedure of random forests is consistent and adapts to sparsity, in the sense that its rate of convergence depends only on the number of strong features and not on how many noise variables are present.

We also mention a current research line: spherical deconvolution by using Lasso-type methods (where we recall that the Lasso is the “canonical” sparse forecaster in the stochastic setting).

5.4. Advances in the application of individual sequences to operational forecasting

Participants: Sébastien Gerchinovitz, Gilles Stoltz.

The results of two MSc internships that took place in 2008 (Sébastien Gerchinovitz) and 2009 (Marie Devaine) were revisited and written up this year in the articles [17] and [23] by Gilles Stoltz and his co-authors.

The first paper is a survey paper that covers the methodology behind the on-line aggregation of predictors for individual sequences and explains the two target applications considered in our project, namely, the prediction of air quality and the forecasting of electricity consumption. It in particular shows that the Lasso is an efficient tool to combine the forecasts of many experts when the number of prediction rounds is small.

The second paper summarizes the empirical results obtained in 2009 during the internship of Marie Devaine at EDF R&D. Its context is to be able to deal efficiently with specialized experts (that only provide predictions in some scenarios, e.g., refrain from predicting during week-ends when they are designed to be efficient for the working days only), while abiding by some operational constraints.

A new collaboration started with economists, see [25]; the mid-term application would be the forecasting of currency exchanges by aggregating the forecasts of some experts (e.g., the ones provided by newspapers).

5.5. On-line aggregation of predictors for the prediction of time series

Participant: Gérard Biau.

Motivated by a broad range of potential applications, we address in [14] the quantile prediction problem of real-valued time series. We present a sequential quantile forecasting model based on the combination of a set of elementary nearest neighbor-type predictors (which play the role of our experts) and show its consistency under a minimum of conditions. Our approach builds on the methodology developed in recent years for prediction of individual sequences and exploits the quantile structure as a minimizer of the so-called pinball loss function. We perform an in-depth analysis of real-world data sets and show that this nonparametric strategy generally outperforms standard quantile prediction methods.

5.6. Calibrated forecasts

Participant: Gilles Stoltz.

In [16] Gilles Stoltz and his co-author provided yet another proof of the existence of calibrated forecasters, with two merits. First, it is valid for an arbitrary finite number of outcomes. Second, it is short and simple and it follows from a direct application of Blackwell's approachability theorem to carefully chosen vector-valued payoff function and convex target set. The proof captures the essence of existing proofs based on approachability (e.g., the proof by Foster in case of binary outcomes) and highlights the intrinsic connection between approachability and calibration. Finally, it leads to the first convergence rates results known in the literature.

6. Contracts and Grants with Industry

6.1. Contracts with Industry

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

7. Other Grants and Activities

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/>);
- Probability and statistics seminar at Université Paris-Sud (Vincent Rivoirard);
- Parisian seminar of statistics at IHP (Vincent Rivoirard; see <https://sites.google.com/site/semstats>).

Grants:

- ANR project in the young researchers track: ATLAS (involves Sébastien Gerchinovitz, Vincent Rivoirard, Gilles Stoltz; see <http://www.math.ens.fr/~stoltz/ANR-ATLAS.html>);
- ANR project in the conception and simulation track: EXPLO/RA (involves Sébastien Gerchinovitz, Gilles Stoltz, Jia Yuan Yu; see <http://sites.google.com/site/anrexplo/ra/>);
- 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. Doctoral studies

Two PhD theses are prepared within our team, by Sébastien Gerchinovitz (2008–present) and Thomas Mainguy (2010–present).

In 2010, three internships at the MSc level took place, all linked to the MSc programme in mathematics at Université Paris-Sud; the students were Thibaut Horel, Clément Levrard, and Thomas Mainguy.

We wrote reports on PhD theses (2 by Olivier Catoni and 2 by Gérard Biau, 2) and on an habilitation (by Gérard Biau) and were examiners for other PhD (8 by Gérard Biau, 1 by Olivier Catoni, 1 by Gilles Stoltz) or habilitation (1 by Gérard Biau) defenses.

8.2. Invited conferences

We only cite the oral communications which we were invited to give at foreign universities and at international conferences or workshops.

Gérard Biau gave talks at the following conferences: “IWAP 2010 (International Workshop on Applied Probability)”, Madrid, Spain, in July 2010, and “Prague Stochastics 2010”, Prague, Czech Republic, in September 2010.

Olivier Catoni gave a talk at the workshop “Foundations and new trends of PAC-Bayesian learning”, London, in March 2010.

Vincent Rivoirard gave a talk at the European Meeting of Statisticians, University of Piraeus, Greece, August 2010, in the invited paper session called “Density estimation by using lasso-type estimators”.

Gilles Stoltz gave a talk at the machine learning seminar at Technion, Haifa, in January 2010.

8.3. Editorial responsibilities

G erard Biau served as an associate editor of the International Statistical Review.

Olivier Catoni is a member of the editorial committee of the joint series of monographies “Math ematiques et Applications” between Springer and SMAI.

Gilles Stoltz was a member of the program committee of the 23rd Conference on Learning Theory (COLT’10); he was awarded the prize of the best reviewer among the members of the program committee.

8.4. Animation of the scientific community

G erard is vice-president of the SFdS (French statistical society), Vincent Rivoirard is a member of the board of the SMAI (French society of applied and industrial mathematics) and its representative to the board of SFdS.

Olivier Catoni was a member of the recruitment committee at INRIA for senior researchers. Vincent Rivoirard was a member of the recruitment committees at Universit  Paris-Sud and Universit  Paris Pierre-et-Marie-Curie. G erard Biau was a member of the recruitment committees at Universit  Paris Pierre-et-Marie-Curie, Universit  Toulouse I, Universit  Paris-Dauphine, and ENSAI Rennes.

G erard Biau was a member of the organization committee of the conference COMPSTAT’10 and of the workshop Journ es du Sud.

G erard Biau is a member of the scientific council of the group of laboratories CREST.

8.5. Teaching

G erard Biau, Vincent Rivoirard, and Gilles Stoltz give series of lectures on their research topics at the MSc (“master 2”) level at Universit  Paris-Sud and Universit  Paris-6.

Olivier Catoni and Gilles Stoltz created –jointly with the INRIA project Sierra (Sylvain Arlot, Jean-Yves Audibert, Francis Bach)– a course at Ecole normale sup rieure, Paris, at the BSc level (“licence 3”) on machine learning.

9. Bibliography

Major publications by the team in recent years

- [1] K. BERTIN, E. LE PENNEC, V. RIVOIRARD. *Adaptive Dantzig density estimation*, in "Annales de l’Institut Henri-Poincar ", 2011, to appear.
- [2] 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.
- [3] 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.
- [4] O. CATONI. *Statistical Learning Theory and Stochastic Optimization — Lectures on Probability Theory and Statistics,  cole d’ t  de Probabilit s de Saint-Flour XXXI – 2001*, Lecture Notes in Mathematics, Springer, 2004, vol. 1851, 269 pages.
- [5] 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>.

- [6] G. LUGOSI, S. MANNOR, G. STOLTZ. *Strategies for prediction under imperfect monitoring*, in "Mathematics of Operations Research", 2008, vol. 33, p. 513–528.
- [7] 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>.
- [8] 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

Articles in International Peer-Reviewed Journal

- [9] G. BIAU, K. BLEAKLEY, L. GYÖRFI, G. OTTUCSÁK. *Nonparametric sequential prediction of time series*, in "Journal of Nonparametric Statistics", 2010, vol. 22, p. 297-317.
- [10] G. BIAU, F. CÉROU, A. GUYADER. *On the rate of convergence of the bagged nearest neighbor estimate*, in "Journal of Machine Learning Research", 2010, vol. 11, p. 687-712.
- [11] G. BIAU, F. CÉROU, A. GUYADER. *Rates of convergence of the functional k -nearest neighbor estimate*, in "IEEE Transactions on Information Theory", 2010, vol. 56, p. 2034-2040.
- [12] G. BIAU, B. CADRE, L. ROUVIÈRE. *Statistical analysis of k -nearest neighbor collaborative recommendation*, in "The Annals of Statistics", 2010, vol. 38, p. 1568-1592.
- [13] G. BIAU, L. DEVROYE. *On the layered nearest neighbour estimate, the bagged nearest neighbour estimate and the random forest method in regression and classification*, in "Journal of Multivariate Analysis", 2010, vol. 101, p. 2499-2518.
- [14] G. BIAU, B. PATRA. *Sequential quantile prediction of time series*, in "IEEE Transactions on Information Theory", 2010, to appear.
- [15] S. BUBECK, R. MUNOS, G. STOLTZ. *Pure Exploration in Finitely-Armed and Continuous-Armed Bandit Problems*, in "Theoretical Computer Science", 2010, to appear.
- [16] S. MANNOR, G. STOLTZ. *A geometric proof of calibration*, in "Mathematics of Operations Research", 2010, to appear.

Articles in National Peer-Reviewed Journal

- [17] G. STOLTZ. *Agrégation séquentielle de prédicteurs : méthodologie générale et applications à la prévision de la qualité de l'air et à celle de la consommation électrique*, in "Journal de la Société Française de Statistique", 2010, vol. 2, n^o 151.

Other Publications

- [18] J.-Y. AUDIBERT, O. CATONI. *Robust linear least squares regression*, 2010, <http://hal.inria.fr/hal-00522534>.
- [19] J.-Y. AUDIBERT, O. CATONI. *Robust linear regression through PAC-Bayesian truncation*, 2010, <http://hal.inria.fr/hal-00522536>.

- [20] G. BIAU. *Analysis of a random forests model*, 2010, <http://hal.inria.fr/hal-00476545>.
- [21] G. BIAU, A. MAS. *PCA-Kernel estimation*, 2010, <http://hal.inria.fr/hal-00467013>.
- [22] O. CATONI. *Challenging the empirical mean and empirical variance: a deviation study*, 2010, <http://hal.inria.fr/hal-00517206>.
- [23] M. DEVAINE, Y. GOUDE, G. STOLTZ. *Forecasting of the electricity consumption by aggregation of specialized experts; application to Slovakian and French country-wide hourly predictions*, 2010, <http://hal.archives-ouvertes.fr/hal-00484940>.
- [24] S. GERCHINOVITZ. *Sparsity regret bounds for individual sequences in online linear regression*, 2010, <http://www.math.ens.fr/~gerchino>.
- [25] T. MICHALSKI, G. STOLTZ. *Do countries falsify economic data strategically? Some evidence that they might do.*, 2010, <http://halshs.archives-ouvertes.fr/halshs-00482106>.