IN PARTNERSHIP WITH:
CNRS

Ecole normale supérieure de Paris

Activity Report 2012

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, Learning and Statistical Methods

Table of contents

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 Project-Team: July 01, 2009 *, Updated into Project-Team:* January 01, 2010 .

1. Members

Research Scientists

Olivier Catoni [Team leader, Senior researcher, CNRS, HdR]
Gilles Stoltz [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; left the team on August 31$^{\text{st}}$]
Ilaria Giulini [PhD student, fellow of Région Ile-de-France]
Thomas Mainguy [PhD student, fellow of Ecole normale supérieure]
Emilien Joly [PhD student, PhD grant of ENS de Cachan]
Paul Baudin [PhD student, fellow of Inria]
Pierre Gaillard [PhD student, PhD student on an industrial contract with EDF R&D]

Administrative Assistant

Lindsay Polienor [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:

- 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;
- natural image analysis, and more precisely the use of unsupervised learning in data representation;
- computational linguistics;
- statistical inference on biological and neurobiological 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 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. 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.

4.5. 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.

4.6. 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 2012 but only published in 2012

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

We do not discuss here the contributions provided by [16], [17], [11], [13], [14], [15], [18], since they were achieved in 2011 or earlier (but only published this year due to the reviewing and publishing process). Also, the book [25] (whose first edition was published in 2009) was augmented and revised for its second edition, published this year.

5.2. Extended journal versions written in 2012 of conference papers published in 2011

Participants: Sébastien Gerchinovitz, Gilles Stoltz.

We wrote extended journal papers of some conference papers discussed in previous annual activity reports; they correspond to refences [32], [19], [20].

5.3. Bayesian methods

Participants: Gérard Biau, Vincent Rivoirard.

5.3.1. The ABC method

Approximate Bayesian Computation (ABC for short) is a family of computational techniques which offer an almost automated solution in situations where evaluation of the posterior likelihood is computationally prohibitive, or whenever suitable likelihoods are not available. In the paper [29] Gérard Biau and his coauthors analyze the procedure from the point of view of k-nearest neighbor theory and explore the statistical properties of its outputs. They discuss in particular some asymptotic features of the genuine conditional density estimate associated with ABC, which is a new interesting hybrid between a k-nearest neighbor and a kernel method.

5.3.2. Semi-parametric version of the Bernstein-von Mises theorem

In [22], Vincent Rivoirard and Judith Rousseau study the asymptotic posterior distribution of linear functionals of the density by deriving general conditions to obtain a semi-parametric version of the Bernstein-von Mises theorem. The special case of the cumulative distributive function evaluated at a specific point is widely considered. In particular, they show that for infinite dimensional exponential families, under quite general assumptions, the asymptotic posterior distribution of the functional can be either Gaussian or a mixture of Gaussian distributions with different centering points. This illustrates the positive but also the negative phenomena that can occur for the study of Bernstein-von Mises results. In [22] Vincent Rivoirard and Judith Rousseau use convergence rates on Besov spaces established in [23].

5.4. Sequential learning

Participants: Pierre Gaillard, Gilles Stoltz.

5.4.1. Bandit problems

The article [30] revisits asymptotically optimal results of Lai and Robbins, Burnetas and Katehakis in a non-asymptotic way. A preliminary attempt was mentioned in the 2011 annual report; it was concerned (essentially) with the case of Bernoulli distributions over the arms. We achieve here the stated optimality of the regret bounds for larger models: regular exponential families; finitely supported distributions.

5.4.2. Theoretical results for the prediction of arbitrary sequences

We generalize and unify in [24] several notions of regret under a same banner: these include adaptive regret (regret against a fixed convex combination on subintervals of the time); shifting regret (regret against a slowly evolving target sequence of convex combinations); and discounted regret (when the instances are weighted with weights depending on how recent the instances are). We recover and sometimes improve some earlier bounds.

5.4.3. Forecasting of the production data of oil reservoirs

We applied our sequential aggregation techniques to a new data set, with IFP Energies nouvelles as a partner. The goal was 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. Results were obtained with the help of an intern, Charles-Pierre Astolfi, and are described in the technical report [27] (to be transformed into a regular journal / conference paper next year).

5.5. Regression, classification, regression methods

Participants: Gérard Biau, Olivier Catoni, Ilaria Giulini.

5.5.1. Metric-based decision procedures

We know now that a good part of the statistical performance of regression and classification algorithms relies on the metric chosen to represent the proximity between the data points. Throughout his work, Gérard Biau became convinced that, well beyond the traditional distances, (dis)similarities and other self-reproducing kernel metrics, it is now necessary to attempt to define proximities generated by the sample itself. These metrics are inevitably random and probabilistic, and force us to rethink the nature of the estimates, as shown for example in the preliminary article [12].

5.5.2. Unsupervised classification in reproducing kernel Hilbert spaces

In her PhD started in September 2012, Ilaria Giulini uses dimension free estimates of the principal components of an i.i.d. sample of points in a Reproducing Kernel Hilbert Space to derive new unsupervised clustering algorithms based on the idea of dimension reduction by nonlinear coordinate smoothing along aggregated principal components. The dimension free estimates are obtained using PAC-Bayes bounds derived from thresholded exponential moments.

5.6. Sparsity and ℓ_1–regularization

Participant: Vincent Rivoirard.

5.6.1. For multivariate Hawkes processes

Motivated by statistical problems in neuroscience, Vincent Rivoirard and his coauthors study in [31] non-parametric inference for multivariate Hawkes processes depending on an unknown function to be estimated by linear combinations of a fixed dictionary. To select coefficients, they propose a Lasso-type methodology where data-driven weights of the penalty are derived from new Bernstein-type inequalities for martingales. Oracle inequalities are established under assumptions on the Gram matrix of the dictionary. Non-asymptotic probabilistic results are proven, which allows them to check these assumptions by considering general dictionaries based on histograms, Fourier or wavelet bases. They finally carry out a simulation study and compare their methodology with the adaptive Lasso procedure proposed by Zou. They observe an excellent behavior of their procedure with respect to the problem of supports recovery. Unlike adaptive Lasso of Zou, their tuning procedure is proven to be robust with respect to all the parameters of the problem, revealing its potential for concrete purposes in neuroscience, but also in other fields.

5.6.2. In the spherical convolution model

In [21], Thanh Mai Pham Ngoc and Vincent Rivoirard consider the problem of estimating a density of probability from indirect data in the spherical convolution model. They aim at building an estimate of the unknown density as a linear combination of functions of an overcomplete dictionary. The procedure is devised through a well-calibrated ℓ_1–penalized criterion. The dictionary approach allows to combine various bases and thus enhances estimates sparsity. They provide an oracle inequality under global coherence assumptions. Moreover, the calibrated procedure that they put forward gives very satisfactory results in the numerical study when compared with other procedures.

5.6.3. For semiparametric nonlinear mixed-effects models

Semiparametric nonlinear mixed-effects models (SNMMs) have been proposed as an extension of nonlinear mixed-effects models (NLMMs). These models are a good compromise and retain nice features of both parametric and nonparametric models resulting in more flexible models than standard parametric NLMMs. In [28], Vincent Rivoirard and his coauthors propose new estimation strategies in SNMMs. They propose a Lasso-type method to estimate the unknown nonlinear function. They derive oracle inequalities for this nonparametric estimator. They combine the two approaches in a general estimation procedure that they illustrate with simulations and through the analysis of a real data set of price evolution in on-line auctions.

5.7. Computational linguistics

Participants: Olivier Catoni, Thomas Mainguy.

In a forthcoming paper, Olivier Catoni and Thomas Mainguy study a new statistical model to learn the syntactic structure of natural languages from a training set made of written sentences. This model learns a new type of stochastic grammar and defines a statistical model on sentences. Global constraints are enforced, that set the approach apart from the family of Markov models. On the other hand, the grammar model generates outputs through a split and merge stochastic process that is more elaborate than the production rules defining a context free grammar. Experiments made on small corpora are very encouraging. Working on large corpora will require to speed up the algorithms used to implement the model as well as some code optimization.

6. Bilateral Contracts and Grants with Industry

6.1. Students Paid by Industrial Partners / Contracts with Industry

- Gérard Biau finished supervising the PhD thesis of Benoît Patra, which took place till March 2012 within an industrial contract (“thèse CIFRE”) with Lokad.com (http://www.lokad.com/).
- Gérard Biau has been supervising the PhD thesis of Baptiste Gregorutti since December 2011, within an industrial contract (“thèse CIFRE”) with Safety Line (http://www.safety-line.fr/index.php/fr/)

- Gilles Stoltz has been supervising the PhD thesis of Pierre Gaillard, which takes place since September 2012 within an industrial contract ("thèse CIFRE") with EDF R&D (http://innovation.edf.com/).
- Gilles Stoltz supervised the M.Sc. internship of Charles-Pierre Astolfi, which took place within a collaboration with IFP Energies nouvelles (http://www.ifpenergiesnouvelles.fr/).

7. Partnerships and Cooperations

7.1. National Initiatives

- ANR project in the conception and simulation track: EXPLO/RA (involves Emilien Joly, Pierre Gaillard, Sébastien Gerchinovitz, Gilles Stoltz; 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/);
- ANR project in the blank program: Calibration (involves Vincent Rivoirard, who is the coordinator; see https://sites.google.com/site/anrcalibration/home);

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, with

- Karine Bertin, University of Valparaiso, Chile;
- Luc Devroye, McGill University, Canada;
- Shie Mannor, Technion, Israel.

In particular, Pierre Gaillard spent 5 months working with Shie Mannor from January to May 2012.

8. Dissemination

8.1. Scientific Animation

8.1.1. Conference organization

Gilles Stoltz was the co-chair of the program committee of the 23rd International Conference on Algorithmic Learning Theory (ALT'12); see the edited volume [26]. He was also a member of the program committee of the 25th Conference on Learning Theory (COLT'12).

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

Olivier Catoni has been a member of the editorial committee of the joint series of monographs “Mathématiques et Applications” between Springer and SMAI until June 2012.

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 a member of the Board of SFdS.

Gérard Biau was elected 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 was elected 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 et Gilles Stoltz, Apprentissage, 20h, niveau L3, Ecole normale supérieure, France

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

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

Master : Gérard Biau, Statistique mathématique, 30h, niveau M1, 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 : Gilles Stoltz, Statistiques et théorie de l’information, 10h, 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, Examinateur à l’oral de probabilités et statistiques de l’agrégation de mathématiques, France

8.2.2. Supervision

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

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, Benoît Patra, Clément Levrard, Benjamin Guedj, Svetlana Gribkova, Baptiste Gregorutti, Erwan Scornet, Nedjmeddine Allab for Gérard Biau, and Laure Sansonnet for Vincent Rivoirard)

MSc theses: Gilles Stoltz supervised the MSc theses of Charles-Pierre Astolfi (MVA, ENS Cachan) and Paul Baudin (MVA, ENS Cachan)

8.2.3. Juries

Gérard Biau was a reviewer for the following PhD defenses:

Ekaterina Sergienko, Université Toulouse III, November 2012

Christophe Denis, Université Paris Descartes, November 2012

Emmanuel Onzon, Université Paris VI, November 2012

and a jury member of the following PhD defenses:

Mohamed Achibi, Université Paris VI, July 2012

Moïse Jérémie, Université Paris VI, September 2012

Virgile Caron, Université Paris VI, October 2012

Caroline Meynet, Université Paris-Sud, November 2012

Nicolas Jégou, Université Rennes 2, November 2012

Sarah Ouadah, Université Paris VI, December 2012

Sylvain Girard, Ecole Nationale Supérieure des Mines de Paris, December 2012

Vincent Rivoirard was a jury member for the following habilitation defenses:

Wintenberger Olivier, Université Paris Dauphine, November 2012

Céline Vial, Université Claude Bernard Lyon 1, Décember 2012

Gérard Biau was a reviewer for the following habilitation defense:

Céline Vial, Université Claude Bernard Lyon 1, Décember 2012

and a jury member of the following habilitation defense:

Fadoua Balabdaoui, Université Paris-Dauphine, May 2012

8.3. Popularization

Gilles Stoltz gave a conference for an audience of students of “classes préparatoires” at Mathematic Park (http://www.ihp.fr/fr/seminaire/mathpark-programme).

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] 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.

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

[9] 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.

[10] 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

[11] G. BIAU. *Analysis of a random forests model*, in "Journal of Machine Learning Research", 2012, vol. 13, p. 1063–1095, http://hal.inria.fr/hal-00476545/en.

[12] G. BIAU, L. DEVROYE, V. DUJMOVIĆ, A. KRZYŻAK. *An affine invariant k-nearest neighbor regression estimate*, in "Journal of Multivariate Analysis", 2012, vol. 112, p. 24–34, http://hal.inria.fr/hal-00655850/.

[13] G. BIAU, A. FISCHER. *Parameter selection for principal curves*, in "IEEE Transactions on Information Theory", 2012, vol. 58, p. 1924–1939, http://hal.inria.fr/hal-00565540/en.

[14] G. BIAU, A. MAS. *PCA-kernel estimation, statistics & risk modeling*, in "Statistics & Risk Modeling", 2012, vol. 29, p. 19–46, http://hal.inria.fr/hal-00467013/en.

[15] G. BIAU, Y. G. YATRACOS. *On the shrinkage estimation of variance and Pitman closeness criterion*, in "Journal de la Société Française de Statistique", 2012, vol. 153, p. 5–21, http://www.lsta.upmc.fr/BIAU/by.pdf.

[16] O. CATONI. *Challenging the empirical mean and empirical variance: a deviation study*, in "Ann. Inst. Henri Poincaré", 2012, vol. 48, n^{o} 4, p. 1148-1185, http://fr.arxiv.org/abs/1009.2048.

[17] M. DEVAINE, P. GAILLARD, Y. GOUDE, G. STOLTZ. *Forecasting electricity consumption by aggregating specialized experts*, in "Machine Learning", 2012, to appear, http://hal.inria.fr/hal-00484940.

[18] M. DOUMIC-JAUFFRET, M. HOFFMANN, P. REYNAUD-BOURET, V. RIVOIRARD. *Nonparametric estimation of the division rate of a size-structured population*, in "SIAM Journal on Numerical Analysis", 2012, vol. 50, n^{o} 2, p. 925–950, http://dx.doi.org/10.1137/110828344.

[19] S. GERCHINOVITZ. *Sparsity regret bounds for individual sequences in online linear regression*, in "Journal of Machine Learning Research", 2012, to appear, http://hal.inria.fr/inria-00552267.

[20] S. GERCHINOVITZ, J. Y. YU. *Adaptive and optimal online linear regression on L1-balls*, in "Theoretical Computer Science", 2012, to appear, http://hal.archives-ouvertes.fr/hal-00594399.

[21] P. NGOC, T. MAI, V. RIVOIRARD. *The dictionary approach for spherical deconvolution*, in "Journal of Multivariate Analysis", 2012, to appear, http://hal.inria.fr/hal-00616519.

[22] V. RIVOIRARD, J. ROUSSEAU. *Bernstein-von Mises theorem for linear functionals of the density*, in "The Annals of Statistics", 2012, vol. 40, n^{o} 3, p. 1489–1523, http://dx.doi.org/10.1214/12-AOS1004.

[23] V. RIVOIRARD, J. ROUSSEAU. *Posterior concentration rates for infinite dimensional exponential families*, in "Bayesian Analysis", 2012, vol. 7, n^{o} 2, p. 311–333, http://hal.archives-ouvertes.fr/hal-00634432/fr/.

International Conferences with Proceedings

[24] N. CESA-BIANCHI, P. GAILLARD, G. LUGOSI, G. STOLTZ. *Mirror descent meets fixed share (and feels no regret)*, in "Advances in Neural Information Processing Systems (NIPS)", Lake Tahoe, United States, December 2012, vol. 25, Paper 471, http://hal.inria.fr/hal-00670514.

Scientific Books (or Scientific Book chapters)

[25] V. RIVOIRARD, G. STOLTZ. *Statistique mathématique en action*, second, Vuibert, 2012, extended and revised version, http://www.dma.ens.fr/statenaction/.

Books or Proceedings Editing

[26] N. BSHOUTY, G. STOLTZ, N. VAYATIS, T. ZEUGMANN (editors). *Proceedings of the 23rd International Conference on Algorithmic Learning Theory (ALT'12)*, Lecture Notes in Computer Science, Springer, 2012, vol. 7568, http://www.springer.com/computer/ai/book/978-3-642-34105-2.

Research Reports

[27] C.-P. ASTOLFI, S. DA VEIGA, G. STOLTZ. *Forecasting production data of oil reservoirs with experts*, IFP Energies nouvelles, 2012, 68 pages.

Other Publications

[28] A. ARRIBAS-GIL, K. BERTIN, C. MEZA, V. RIVOIRARD. *Lasso-type estimators for semiparametric nonlinear mixed-effects models estimation*, 2012, http://hal.inria.fr/hal-00665843.

[29] G. BIAU, F. CÉROU, A. GUYADER. *New Insights into Approximate Bayesian Computation*, 2012, http://hal.inria.fr/hal-00721164.

[30] O. CAPPÉ, A. GARIVIER, O.-A. MAILLARD, R. MUNOS, G. STOLTZ. *Kullback-Leibler Upper Confidence Bounds for Optimal Sequential Allocation*, 2012, http://hal.inria.fr/hal-00738209.

[31] N.-R. HANSEN, P. REYNAUD-BOURET, V. RIVOIRARD. *Lasso and probabilistic inequalities for multivariate point processes*, 2012, http://hal.inria.fr/hal-00722668.

[32] S. MANNOR, V. PERCHET, G. STOLTZ. *Robust approachability and regret minimization in games with partial monitoring*, 2012, http://hal.inria.fr/hal-00595695.