PAC-Bayesian aggregation and multi-armed bandits

Jean-Yves Audibert

PAC-Bayesian aggregation and multi-armed bandits

(Habilitation thesis)

Jean-Yves AUDIBERT

Chapter 1

Introduction

This document presents the research I have undertaken since the beginning of my PhD thesis. The Laboratoire de Probabilits et Modles Alatoires of Universit Paris 6 hosted my PhD (2001-2004). I was then recruited in the Centre d'Enseignement et de Recherche en Traitement de l'Information et Signal de l'Ecole Nationale des Ponts et Chausses, which is now a common research laboratory with the Centre Scientifique et Technique du Btiment and part of the Laboratoire d'informatique Gaspard Monge de l'Universit Paris Est. Besides, since 2007, part of my research has been done within the Willow team of the Laboratoire d'Informatique de l'cole Normale Superiure.

My main research directions are statistical learning theory and machine learning techniques for computer vision. Machine learning is a research field positioned between statistics, computer science and applied mathematics. Its goal is to bring out theories and algorithms to better understand and deal with complex systems for which no simple, accurate and easy-to-use model exists. It has a considerable impact on a wide variety of scientific domains, including text analysis and indexation, financial market analysis, search engines, bioinformatics, speech recognition, robotics, industrial engineering... The development of new sensors to acquire data, the increasing capacity of storage and computational power of computers have brought new perspectives to understand more and more complex systems from observations. In particular, Machine learning techniques are used in computer vision tasks that are unsolvable using classical methods (object detection, handwriting recognition, image segmentation and annotation).

The core problem in statistical learning can be formalized in the following way. We observe n input-output (or object-label) pairs: $Z_1 = (X_1, Y_1), \ldots, Z_n = (X_n, Y_n)$. A new input X comes. The goal is to predict its associated output Y. The input is usually high dimensional and highly structured (such as a digital image). The output is simple: it is typically a real number or an element in a finite set (for instance, 'yes' or 'no' in the case of the detection of a specific object in the digital image). The usual probabilistic modelling is that the observed data (or training set) and the input-output pair $Z = (X, Y)$ are independent and identically distributed random variables coming from some unknown distribution P, and that, for various possible reasons, the output is not necessarily a deterministic function of the input.

The lack of quality of a prediction y' when y is the true output is measured by its loss, denoted $\ell(y, y')$. Typical loss functions are the 0/1 loss: $\ell(y, y') = \mathbb{1}_{y \neq y'}$ (the loss is one if and only if the prediction differs from the true output) and the square loss: $\ell(y, y') = (y - y')^2$. The latter loss is more appropriate than the 0/1 loss when the output space is the real line, a small difference between the

prediction and the true output generating a small loss. The target of learning is to infer from the training set a function g from the input space to the output space having a low risk, also called expected loss or generalization error:

$$R(g) = \mathbb{E}_{(X,Y)\sim P}\ell\big(Y, g(X)\big).$$

Statistical learning theory aims at answering the following questions. What are the conditions for (asymptotic) consistency of the learning scheme? What can we learn from a finite sample of observations? Under which circumstances, can we expect the risk to be close to the risk of the best prediction function, that is the one we could have proposed had we a full knowledge of the probability distribution P underlying the observations? How accurate is the prediction built on the training set? For instance, how low is its risk? What kind of guarantees can we ensure? Both theoretical and empirical (i.e., computable from the observed data) upper bounds on the risk or the excess risk are of interest. Can we understand/explain the success of some prediction schemes? Besides, we also expect that a new theoretical analysis leads to the design of new prediction methods.

This document details my contributions to these issues, and specifically to:

- the PAC-Bayesian analysis of statistical learning,
- the three aggregation problems: given d functions, how to predict as well as
  - the best of these d functions (model selection type aggregation),
  - the best convex combination of these d functions,
  - the best linear combination of these d functions,
- the multi-armed bandit problems,

Being in computer science departments where image processing and computer vision are core research directions leads me to address a wide variety of topics in which machine learning plays a key role. It includes object recognition, content-based image retrieval, image segmentation and image annotation and vanishing point detection. This document will not detail my contributions on these topics. My related publications can be found on my webpage.

Chapter 2

The PAC-Bayesian analysis of statistical learning

2.1. INTRODUCTION

The natural target of learning is to predict as well as if we had known the distribution generating the input-output pairs. In other words, we want to infer from the training set $Z_1^n = \{(X_1, Y_1), \ldots, (X_n, Y_n)\}$ a prediction function $\hat{g}$ whose risk is close to the risk of the Bayes predictor $g^* = \text{argmin}_g R(g)$, where the minimum is taken among all functions $g : \mathcal{X} \to \mathcal{Y}$ (such that $\ell\big(Y, g(X)\big)$ is integrable). The goal is therefore to propose a good estimator $\hat{g}$ of g^*, where the quality of the estimator is not in terms of the functional proximity of the prediction functions but in terms of their risk similarity.

Since the distribution P of the input-output pair is unknown, the risk is not observed, and numerous core learning procedures have recourse to its empirical counterpart:

$$r(g) = \frac{1}{n} \sum_{i=1}^{n} \ell(Y_i, g(X_i)),$$

either by minimizing it on a restricted class of functions, or almost equivalently by minimizing a linear combination of this empirical risk and a penalty (or regularization) term whose role is to favor "simple" functions. The term "simple" typically refers to some a priori of the statistician, and is often linked to either some smoothness property or some sparsity of the prediction function. The traditional approach to statistical learning theory relies on the study of $R(\hat{g}) - r(\hat{g})$.

In the PAC-Bayesian approach, randomized prediction schemes are considered. Let $\mathcal{M}$ denote the set of distributions on the set $\mathcal{G}(\mathcal{X}; \mathcal{Y})$ of functions from the input space to the output space. A distribution $\hat{\rho}$ in $\mathcal{M}$ is chosen from the data, and the quantity of interest is $R(g)$, where g is drawn from the distribution $\hat{\rho}$. This risk is thus doubly stochastic: it depends on the realization of the training set (which is a realization of the n-fold product distribution $P^{\otimes n}$ of P) and on the realization of the (posterior) distribution $\hat{\rho}$.

Basically, one can argue that the difference between the approaches seems minor: the understanding of $\mathbb{E}_{g\sim\hat{\rho}} R(g)$ for any distribution $\hat{\rho}$ implies the understanding of $R(\hat{g})$ (simply by considering the Dirac distribution at $\hat{g}$), and that the converse is also true (to the extent that if $R(\hat{g}) \leq B(\hat{g})$ holds for any estimator $\hat{g}$ and some real-valued function B, then $\mathbb{E}_{g\sim\hat{\rho}} R(g) \leq \mathbb{E}_{g\sim\hat{\rho}} B(g)$ also holds for any posterior distribution $\hat{\rho}$).

The main difference lies rather in the very starting point of the PAC-Bayesian analysis. To detail it, let me introduce a distribution $\pi \in \mathcal{M}$, that is non-random (as opposed to $\hat{\rho}$, which depends on the sample). The central argument is (based on)

the following property of the Kullback-Leibler (KL) divergence: for any bounded function $h : \mathcal{G}(\mathcal{X}; \mathcal{Y}) \to \mathbb{R}$, we have

$$\sup_{\rho \in \mathcal{M}} \left\{ \mathbb{E}_{g \sim \rho} h(g) - K(\rho, \pi) \right\} = \log \mathbb{E}_{g \sim \pi} e^{h(g)}, \tag{2.1.1}$$

where e denotes the exponential number, and $K(\rho, \pi)$ is the KL divergence between the distributions ρ and π: $K(\rho, \pi) = \mathbb{E}_{g \sim \rho} \log \left(\frac{\rho}{\pi}(g) \right)$ if ρ admits a density with respect to π, denoted $\frac{\rho}{\pi}$, and $K(\rho, \pi) = +\infty$ otherwise[1]. To control the difference $\mathbb{E}_{g \sim \hat{\rho}} R(g) - \mathbb{E}_{g \sim \hat{\rho}} r(g)$, putting aside integrability issues, one essentially uses: for any $\lambda > 0$,

$$\begin{aligned}
\mathbb{E}_{Z_1^n \sim P^{\otimes n}} e^{\lambda [\mathbb{E}_{g \sim \hat{\rho}} R(g) - \mathbb{E}_{g \sim \hat{\rho}} r(g)] - K(\hat{\rho}, \pi)} &\leq \mathbb{E}_{Z_1^n \sim P^{\otimes n}} e^{\sup_{\rho \in \mathcal{M}} \lambda [\mathbb{E}_{g \sim \rho} R(g) - \mathbb{E}_{g \sim \rho} r(g)] - K(\rho, \pi)} \\
&= \mathbb{E}_{Z_1^n \sim P^{\otimes n}} \mathbb{E}_{g \sim \pi} e^{\lambda [R(g) - r(g)]} \\
&= \mathbb{E}_{g \sim \pi} \mathbb{E}_{Z_1^n \sim P^{\otimes n}} e^{\lambda [R(g) - r(g)]} \\
&= \mathbb{E}_{g \sim \pi} \left(\mathbb{E}_{(X,Y) \sim P} e^{\frac{\lambda}{n} [R(g) - \ell(Y, g(X))]} \right)^n .
\end{aligned} \tag{2.1.2}$$

A first consequence is that PAC-Bayes bounds are not (directly) useful for posterior distributions with $K(\hat{\rho}, \pi) = +\infty$: this is in particular the case when $\hat{\rho}$ is a Dirac distribution and π assigns no probability mass to single functions. So classical results of the standard approach does not derive from the PAC Bayesian approach. On the other hand, the apparition of the KL term shows that the PAC-Bayesian analysis fundamentally differs from the simple analysis given in the previous paragraph.

To illustrate this last point, consider the case of a prior distribution putting mass on a finite set $\mathcal{G} \subset \mathcal{G}(\mathcal{X}; \mathcal{Y})$ of functions. For simplicity, consider bounded losses, say $0 \leq \ell(y, y') \leq 1$ for any $y, y' \in \mathcal{Y}$. By using Hoeffding's inequality and a weighted union bound, one gets that for any $\varepsilon > 0$, with probability at least $1 - \varepsilon$, we have for any $g \in \mathcal{G}$

$$R(g) - r(g) \leq \sqrt{\frac{\log(\pi^{-1}(g) \varepsilon^{-1})}{2n}},$$

hence for any distribution ρ such that $\rho(\mathcal{G}) = 1$,

$$\begin{aligned}
\mathbb{E}_{g \sim \rho} R(g) - \mathbb{E}_{g \sim \rho} r(g) &\leq \mathbb{E}_{g \sim \rho} \sqrt{\frac{\log(\pi^{-1}(g) \varepsilon^{-1})}{2n}} \\
&\leq \sqrt{\frac{K(\rho, \pi) + H(\rho) + \log(\varepsilon^{-1})}{2n}},
\end{aligned} \tag{2.1.3}$$

[1] See Appendix A for a summary of the properties of the KL divergence.

where the second inequality uses Jensen's inequality and Shannon's entropy: $H(\rho) = -\sum_{g\in\mathcal{G}} \rho(g)\log\rho(g)$. This is to be compared to the first PAC-Bayesian bound from the pioneering work of McAllester [102], which states that with probability at least $1-\varepsilon$, for any distribution $\rho \in \mathcal{M}$, we have

$$\mathbb{E}_{g\sim\rho} R(g) - \mathbb{E}_{g\sim\rho} r(g) \leq \sqrt{\frac{K(\rho,\pi) + \log(n) + 2 + \log(\varepsilon^{-1})}{2n-1}}.$$

The main difference is that the Shannon entropy has been replaced with a $\log n$ term. In fact, the latter bound is not restricted to prior distributions putting mass on a finite set of functions: it is valid for any distribution π. On the contrary, the basic argument leading to (2.1.3) does not extend to continuous set of functions because of the Shannon's entropy term (for ρ putting masses on a continuous set of functions, this term diverges).

The previous discussion has shown the originality of the PAC-Bayesian analysis. However it does not clearly demonstrate its usefulness. Several works in the last decade have shown that the approach is indeed useful, and that PAC-Bayesian bounds lead to tight bounds, which are often representative of the risk behaviour even for relatively small training sets (see e.g. [88, 103, 82] for margin-based bounds from Gaussian prior distributions, [83] for an Adaboost setting, that is majority vote of weak learners, [118] in a clustering setting, [7, Chap.2],[89] for compression schemes, [50, 51] for PAC bounds with sparsity-inducing prior distributions).

My contributions to the PAC-Bayesian approach are the use of relative PAC-Bayesian bounds to design estimators with minimax rates (Section 2.3), the combination of the PAC-Bayesian argument with metric and (generic) chaining arguments (Section 2.4), the use of PAC-Bayesian bounds to propose new estimators and minimax bounds under weak assumptions for the aggregation problems (Chapter 3). Before detailing them, I give in the next section a global picture of PAC-Bayesian bounds, with a particular emphasis on the relations between the different works since they have not been underlined so far in the literature.

2.2. PAC-BAYESIAN BOUNDS

We consider that the losses are between 0 and 1, unless otherwise stated. The symbol C will be used to denote a constant that may differ from line to line. The bounds stated here are the original ones, possibly up to minor improvements. Most of them rely on a different use of the duality formula (2.1.1) and the Markov inequality, which allows to prove a Probably Approximatively Correct (PAC) bound from the control of the Laplace transform of an appropriate random variable: pre-

cisely, if a real-valued random variable V is such that $\mathbb{E}e^V \leq 1$, then for any $\varepsilon > 0$, with probability at least $1-\varepsilon$, $V \leq \log(\varepsilon^{-1})$.

2.2.1. NON LOCALIZED PAC-BAYESIAN BOUNDS. McAllester's first bound states that for any $\varepsilon > 0$, with probability at least $1-\varepsilon$, for any $\rho \in \mathcal{M}$, we have

$$\mathbb{E}_{g\sim\rho}R(g) - \mathbb{E}_{g\sim\rho}r(g) \leq \sqrt{\frac{K(\rho,\pi) + \log(2n) + \log(\varepsilon^{-1})}{2n-1}}. \tag{McA}$$

In [87, 117], Seeger has proposed a simplified proof and improved the bound when the losses take only two values 0 or 1 (classification losses). The result is that with probability at least $1-\varepsilon$, for any $\rho \in \mathcal{M}$, we have

$$K(\mathbb{E}_{g\sim\rho}r(g), \mathbb{E}_{g\sim\rho}R(g)) \leq \frac{K(\rho,\pi) + \log(2\sqrt{n}\varepsilon^{-1})}{n}. \tag{S}$$

where, with a slight abuse of notation, $K(\mathbb{E}_{g\sim\rho}r(g), \mathbb{E}_{g\sim\rho}R(g))$ denotes the KL divergence between the Bernoulli distributions of respective parameters $\mathbb{E}_{g\sim\rho}r(g)$ and $\mathbb{E}_{g\sim\rho}R(g)$. The concise proofs of (McA) and (S) are given in Appendices B and C.

Since we have $\mathbb{E}_{g\sim\rho}R(g) - \mathbb{E}_{g\sim\rho}r(g) \leq \sqrt{K\big(\mathbb{E}_{g\sim\rho}r(g), \mathbb{E}_{g\sim\rho}R(g)\big)}$ (Pinsker's inequality), (S) implies (McA). Besides, when $\mathbb{E}_{g\sim\rho}R(g)$ is small, (S) provides a much better bound than (McA) since, from a cumbersome study of the function $t \mapsto K(\mathbb{E}_{g\sim\rho}r(g), \mathbb{E}_{g\sim\rho}r(g)+t)$, (S) implies

$$\big|\mathbb{E}_{g\sim\rho}R(g) - \mathbb{E}_{g\sim\rho}r(g)\big| \leq \sqrt{\frac{2\mathbb{E}_{g\sim\rho}r(g)[1-\mathbb{E}_{g\sim\rho}r(g)]\mathcal{K}}{n}} + \frac{4\mathcal{K}}{3n}, \tag{S'}$$

with $\mathcal{K} = K(\rho,\pi) + \log(2\sqrt{n}\varepsilon^{-1})$. In particular, when the empirical risk of the randomized estimator is zero, this last bound is of $1/n$ order, while (McA) only gives a $1/\sqrt{n}$ order.

Still in the classification setting, Catoni [40] proposed a different bound: for any $\varepsilon > 0$ and $\lambda > 0$ with $\frac{\lambda}{n}\Psi(\frac{\lambda}{n}) < 1$, with probability at least $1-\varepsilon$, for any $\rho \in \mathcal{M}$,

$$\mathbb{E}_{g\sim\rho}R(g) \leq \frac{\mathbb{E}_{g\sim\rho}r(g)}{1-\frac{\lambda}{n}\Psi(\frac{\lambda}{n})} + \frac{K(\rho,\pi)+\log(\varepsilon^{-1})}{\lambda[1-\frac{\lambda}{n}\Psi(\frac{\lambda}{n})]}, \tag{C1}$$

where

$$\Psi(t) = \frac{e^t - 1 - t}{t^2}.$$

Since typical values of λ (the ones which minimizes the previous right-hand side) are in $[C\sqrt{n}; Cn]$ and since $\Psi(\lambda/n) \approx 1/2$ for λ/n close to 0, we roughly have

$$\mathbb{E}_{g\sim\rho}R(g) \lessapprox \mathbb{E}_{g\sim\rho}r(g) + \frac{\lambda}{2n}\mathbb{E}_{g\sim\rho}r(g) + \frac{K(\rho,\pi)+\log(\varepsilon^{-1})}{\lambda},$$

which gives by choosing λ optimally[2]

$$\mathbb{E}_{g\sim\rho}R(g) \lessapprox \mathbb{E}_{g\sim\rho}r(g) + \sqrt{2\mathbb{E}_{g\sim\rho}r(g)\frac{K(\rho,\pi)+\log(\varepsilon^{-1})}{n}}. \qquad \text{(C1')}$$

My PhD thesis used in variant ways the following Bernstein's type PAC-Bayesian bound, which is a direct extension of the argument giving (C1): for any $\lambda > 0$, with probability at least $1-\varepsilon$, for any $\rho \in \mathcal{M}$,

$$\mathbb{E}_{g\sim\rho}R(g) \le \mathbb{E}_{g\sim\rho}r(g) + \frac{\lambda}{n}\Psi\left(\frac{\lambda}{n}\right)\mathbb{E}_{g\sim\rho}\,\mathbb{V}\mathrm{ar}_Z\,\ell(Y,g(X)) + \frac{K(\rho,\pi)+\log(\varepsilon^{-1})}{\lambda}. \qquad \text{(A)}$$

The basic PAC-Bayesian bound used in Zhang's works [136, 137] does not require any boundedness assumption of the loss function and states that for any $\lambda > 0$, with probability at least $1-\varepsilon$, for any $\rho \in \mathcal{M}$,

$$-\frac{n}{\lambda}\mathbb{E}_{g\sim\rho}\log\mathbb{E}_Z e^{-\frac{\lambda}{n}\ell(Y,g(X))} \le \mathbb{E}_{g\sim\rho}r(g) + \frac{K(\rho,\pi)+\log(\varepsilon^{-1})}{\lambda}. \qquad \text{(Z)}$$

Catoni's book [41] concentrates on the classification task. Instead of using

$$\log\mathbb{E}e^{-\frac{\lambda}{n}\ell(Y,g(X))} \le -\frac{\lambda}{n}R(g) + \frac{\lambda^2}{n^2}\Psi\Big(\frac{\lambda}{n}\Big)R(g),$$

which would give (C1) from (Z), Catoni used the equality

$$\log\mathbb{E}e^{-\frac{\lambda}{n}\ell(Y,g(X))} = \log\big(1 - R(g)(1-e^{-\frac{\lambda}{n}})\big),$$

and obtain that with probability at least $1-\varepsilon$, for any $\rho \in \mathcal{M}$,

$$-\frac{n}{\lambda}\log[1-(1-e^{-\frac{\lambda}{n}})\mathbb{E}_{g\sim\rho}R(g)] \le \mathbb{E}_{g\sim\rho}r(g) + \frac{K(\rho,\pi)+\log(\varepsilon^{-1})}{\lambda}. \qquad \text{(C2)}$$

To compare Seeger's bound with the bounds having the free parameter λ in the classification framework, one needs to apply the same kind of analysis which leads from (C1) to (C1'). As a result, both (A) and (Z) lead to

$$\mathbb{E}_{g\sim\rho}R(g) \lessapprox \mathbb{E}_{g\sim\rho}r(g) + \sqrt{2\mathbb{E}_{g\sim\rho}\big(R(g)[1-R(g)]\big)\frac{K(\rho,\pi)+\log(\varepsilon^{-1})}{n}}, \qquad \text{(Z')}$$

[2]Technically speaking, we are not allowed to choose λ depending on ρ, but using a union bound argument, the argument can be made rigorous at the price that the $\log(\varepsilon^{-1})$ term becomes $\log(C\log(Cn)\varepsilon^{-1})$.

(C1) leads to

$$\mathbb{E}_{g\sim\rho}R(g) \lessapprox \mathbb{E}_{g\sim\rho}r(g) + \sqrt{2\mathbb{E}_{g\sim\rho}R(g)\frac{K(\rho,\pi)+\log(\varepsilon^{-1})}{n}},$$

(S) gives, once more, from studying the function $t \mapsto K(\mathbb{E}_{g\sim\rho}r(g), \mathbb{E}_{g\sim\rho}r(g)+t)$,

$$\mathbb{E}_{g\sim\rho}R(g) \leq \mathbb{E}_{g\sim\rho}r(g) + \sqrt{\frac{2\mathbb{E}_{g\sim\rho}R(g)[1-\mathbb{E}_{g\sim\rho}R(g)]\mathcal{K}}{n}} + \frac{2\mathcal{K}}{3n},$$

with $\mathcal{K} = K(\rho,\pi) + \log(2\sqrt{n}\varepsilon^{-1})$, and finally (C2) leads to

$$\mathbb{E}_{g\sim\rho}R(g) \lessapprox \mathbb{E}_{g\sim\rho}r(g) + \sqrt{2\big(\mathbb{E}_{g\sim\rho}R(g)[1-\mathbb{E}_{g\sim\rho}R(g)]\big)\frac{K(\rho,\pi)+\log(\varepsilon^{-1})}{n}}.$$

Although we have $\mathbb{E}_{g\sim\rho}R(g)[1-\mathbb{E}_{g\sim\rho}R(g)] \geq \mathbb{E}_{g\sim\rho}R(g)[1-R(g)]$ (from Jensen's inequality), the two quantities will be of the same order, and also of the order of $\mathbb{E}_{g\sim\rho}R(g)$ for the typical posterior distributions, i.e., the ones which concentrate on low risk functions. As a consequence, in the classification setting, all these bounds are similar (even if this similarity has not been exhibited so far in the literature).

In fact, the works which lead to (C1), (A), (Z) and (C2) rather differ in the way these bounds are refined and used. The main common refinement is the PAC-Bayesian localization, which can be seen as a way to reduce the complexity term and the influence of the particular choice of the prior distribution π. Before detailing the localization idea, let us see how to design an estimator from PAC-Bayesian bounds.

2.2.2. FROM PAC-BAYESIAN BOUNDS TO ESTIMATORS. The standard way to exploit an upper bound on the risk of any estimators is to minimize it in order to get the estimator having the best guarantee in view of the bound. This will be achievable if the bound is empirical, that is computable from the observations. Bounds (McA), (S'), (C1), (A) and (C2) are of this type (unlike (Z') for instance).

When minimizing PAC-Bayesian bounds, one gets a posterior distribution corresponding to a randomized estimator. The minimizer can be written in the following form

$$\pi_h(dg) = \frac{e^{h(g)}}{\mathbb{E}_{g'\sim\pi}e^{h(g')}} \cdot \pi(dg)$$

for some appropriate function $h : \mathcal{G} \to \mathbb{R}$. This is essentially due to the equality $\operatorname{argmin}_{\rho\in\mathcal{M}}\big\{ -\mathbb{E}_{g\sim\rho}h(g) + K(\rho,\pi)\big\} = \pi_h$.

Let us now detail the case of McAllester's bound as it is representative of what can be derived from the other PAC-Bayesian bounds. Let $B(\rho) = \mathbb{E}_{g\sim\rho}r(g) +$

$\sqrt{\frac{K(\rho,\pi)+\log(4n\varepsilon^{-1})}{2n-1}}$. McAllester's bound implies that for any distribution $\rho \in \mathcal{M}$, we have $\mathbb{E}_{g\sim\rho}R(g) \leq B(\rho)$. From this, one can deduce that there exists $\hat{\lambda} \in [\lambda_1, \lambda_2]$ s.t. $B(\pi_{-\hat{\lambda}r}) = \min_\rho B(\rho)$ with $\lambda_1 = \sqrt{4(2n-1)\log(4n\varepsilon^{-1})}$ and $\lambda_2 = 2\lambda_1 + 4(2n-1)$. Besides, the parameter $\hat{\lambda}$ which can be called inverse temperature parameter by analogy with the Boltzmann distribution in statistical mechanics satisfies

$$\hat{\lambda} = \sqrt{4(2n-1)[K(\pi_{-\hat{\lambda}r}, \pi) + \log(4n\varepsilon^{-1})]}$$

and $\hat{\lambda} \in \operatorname*{argmin}_{\lambda>0} \Big\{ -\frac{1}{\lambda}\log \mathbb{E}_{g\sim\pi} e^{-\lambda r(g)} + \frac{\lambda}{4(2n-1)} + \frac{\log(4n\varepsilon^{-1})}{\lambda} \Big\}$.

The posterior distribution is thus a distribution which concentrates on low empirical risk functions, but is still a bit diffuse since to avoid a high KL complexity term, the optimal parameter $\hat{\lambda}$ cannot be larger than Cn. The next section shows how to reduce the complexity term by tuning the prior distribution.

2.2.3. LOCALIZED PAC-BAYESIAN BOUNDS. Without prior knowledge, one may want to choose a prior distribution π which is rather "flat". Now for a particular choice of posterior distribution $\hat{\rho}$, from the equality $\mathbb{E}_{Z_1^n} K(\hat{\rho}, \pi) = \mathbb{E}_{Z_1^n} K(\hat{\rho}, \mathbb{E}_{Z_1^n}[\hat{\rho}]) + K(\mathbb{E}_{Z_1^n}[\hat{\rho}], \pi)$, the prior distribution (recall that it is not allowed to depend on the training set) which minimize the expectation of the KL divergence is $\mathbb{E}_{Z_1^n}\hat{\rho}$, where the expectation is taken with respect to the training set distribution[3]. Now using such a prior distribution does not lead to empirical bound. To alleviate this issue and since the typical posterior distributions have the form $\pi_{-\lambda r}$ for some $\lambda > 0$ (as seen in the previous section), one may consider the prior distribution $\pi_{-\beta R}$ for some $\beta > 0$, use the expansion

$$K(\rho, \pi_{-\beta R}) = K(\rho, \pi) + \beta \mathbb{E}_{g\sim\rho} R(g) + \log\big(\mathbb{E}_{g\sim\pi} e^{-\beta R(g)}\big),$$

and obtain an empirical bound by controlling the last non-observable term by its empirical version.

This leads to the following localized PAC-Bayesian bound which was obtained by Catoni in [40]: for any $\varepsilon > 0$, $\lambda > 0$ and $\xi \geq 0$ such that $\frac{(1+\xi)\lambda}{(1-\xi)n}\Psi(\frac{\lambda}{n}) < 1$, with probability at least $1-\varepsilon$, for any $\rho \in \mathcal{M}$, we have

$$\mathbb{E}_{g\sim\rho} R(g) \leq \frac{\mathbb{E}_{g\sim\rho} r(g)}{1 - \frac{(1+\xi)\lambda}{(1-\xi)n}\Psi(\frac{\lambda}{n})} + \frac{K(\rho, \pi_{-\xi\lambda r}) + (1+\xi)\log(2\varepsilon^{-1})}{(1-\xi)\lambda[1 - \frac{(1+\xi)\lambda}{(1-\xi)n}\Psi(\frac{\lambda}{n})]}. \qquad \text{(C3)}$$

[3] As noted by Catoni, $\mathbb{E}_{Z_1^n} K(\hat{\rho}, \mathbb{E}_{Z_1^n}[\hat{\rho}])$ is exactly the mutual information of the random variable $\hat{g}$ drawn according to the posterior distribution $\hat{\rho}$ and the training sample Z_1^n. This makes a nice connexion between the learning rate of a randomized estimator and information theory.

The parameter ξ characterizes the localization. For $\xi = 0$, we recover (C1) (up to a minor difference on the confidence level). For $\xi > 0$, the KL term is (potentially much) smaller when considering the posterior distribution $\pi_{-\gamma r}$ with $\gamma \geq \xi\lambda$.

We use similar ideas in the case of the comparison of the risks of two randomized estimators as we will see in Section 2.3. Zhang [136, 137] localizes by using π_h with $h(g) = \alpha \log \mathbb{E}_Z e^{-\lambda \ell(Y, g(X))}$ instead of $\pi_{-\beta R}$. The argument there is slightly different and does not lead to empirical bounds on the risk of the randomized estimator with posterior distribution of the form $\pi_{-\lambda r}$. Nevertheless, it was sufficient to prove tight theoretical bounds for this estimator in different contexts: density estimation, classification and least squares regression.

Ambroladze, Parrado-Hernández and Shawe-Taylor [6] proposed a different way to reduce the influence of a "flat" prior distribution. Their localization scheme is based on cutting the training set into two parts and learn from the first part the prior distribution to be used on the second part of the training set. Catoni [41] uses $\pi_{-n \log[1+(e^{\beta/n}-1)R]}$ to obtain tighter localized bounds in the classification setting. Alquier [4, 5] uses $\pi_{-\beta R}$ for general unbounded losses with application to regression and density estimation.

2.3. Comparison of the risk of two randomized estimators

2.3.1. Relative PAC-Bayesian bounds. My PhD (its second chapter) used relative bounds which compare the risk of two randomized estimators to design new (randomized) estimators. The rationale behind developing this type of bounds is that the fluctuations of $R(g_2) - R(g_1) + r(g_1) - r(g_2)$ can be much smaller than the fluctuations of $R(g_2) - r(g_2)$, and this can lead to significantly tighter bounds. Technically speaking, relative bounds are deduced from standard bounds by replacing $\mathcal{G}$ by $\mathcal{G} \times \mathcal{G}$, taking the loss $\ell(y, (g_1, g_2)(x)) = \ell(y, g_2(x)) - \ell(y, g_1(x))$ (with a slight abuse of notation) and by considering product distributions on $\mathcal{G} \times \mathcal{G}$, i.e. $\rho = \rho_1 \otimes \rho_2$ with ρ_1 and ρ_2 distributions on $\mathcal{G}(\mathcal{X}; \mathcal{Y})$. This standard argument transforms (A) into the following assertion holding for losses taking values in $[0, 1]$. For any $\lambda > 0$ and (prior) distributions π_1 and π_2 in $\mathcal{M}$, with probability at least $1 - \varepsilon$, for any $\rho_1 \in \mathcal{M}$ and $\rho_2 \in \mathcal{M}$,

$$\begin{aligned}\mathbb{E}_{g_2\sim\rho_2} R(g_2) - \mathbb{E}_{g_1\sim\rho_1} R(g_1) \leq{}& \mathbb{E}_{g_2\sim\rho_2} r(g_2) - \mathbb{E}_{g_1\sim\rho_1} r(g_1) \\ &+ \frac{\lambda}{n} \Psi\left(\frac{\lambda}{n}\right) \mathbb{E}_{g_2\sim\rho_2} \mathbb{E}_{g_1\sim\rho_1} \mathbb{E}_Z \left([\ell(Y, g_1(X)) - \ell(Y, g_2(X))]^2 \right) \\ &+ \frac{K(\rho_2, \pi_2) + K(\rho_1, \pi_1) + \log(\varepsilon^{-1})}{\lambda}. \end{aligned} \tag{2.3.1}$$

Getting empirical relative bounds calls for controlling the variance term. This is achieved by plugging the following inequality, which holds with probability at

least $1-\varepsilon$, into the previous one

$$\mathbb{E}_{g_2\sim\rho_2}\mathbb{E}_{g_1\sim\rho_1}\mathbb{E}_Z[\ell(Y,g_1(X))-\ell(Y,g_2(X))]^2 \leq \left(1+\frac{\lambda}{2n}\right)\mathbb{E}_{g_2\sim\rho_2}\mathbb{E}_{g_1\sim\rho_1}\frac{1}{n}\sum_{i=1}^n[\ell(Y_i,g_1(X_i))-\ell(Y_i,g_2(X_i))]^2 + \left(1+\frac{\lambda}{2n}\right)^2\frac{K(\rho_2,\pi_2)+K(\rho_1,\pi_1)+\log(\varepsilon^{-1})}{\lambda}.$$

Now, the localization argument described in Section 2.2.3 no longer works as it would change the left-hand side of (2.3.1) into $(1+\xi_2)\mathbb{E}_{g_2\sim\rho_2}R(g_2)-(1-\xi_1)\mathbb{E}_{g_1\sim\rho_1}R(g_1)$ for some positive constants ξ_1 and ξ_2, and would therefore fail to produce relative bounds. To solve this issue, I proved the following uniform empirical upper bound on the KL divergence with respect to a localized prior: for any $\epsilon>0$ and $0<\lambda\leq 0.19\,n$, with probability at least $1-2\varepsilon$, for any $\rho\in\mathcal{M}$, we have

$$K\big(\rho,\pi_{-\lambda R}\big)\leq 2K\big(\rho,\pi_{-\lambda r}\big) + 2\log\mathbb{E}_{g_1\sim\pi_{-\lambda r}}e^{\frac{4\lambda^2}{n}\mathbb{E}_{g_2\sim\rho}\frac{1}{n}\sum_{i=1}^n[\ell(Y_i,g_1(X_i))-\ell(Y_i,g_2(X_i))]^2}+\log(\varepsilon^{-1}),$$

and get the following localized empirical PAC-Bayesian relative bound: for any $\lambda>0$ and $0<\lambda_1,\lambda_2\leq 0.19\,n$, with probability at least $1-\varepsilon$,

$$\begin{aligned}\mathbb{E}_{g_2\sim\rho_2}R(g_2)-\mathbb{E}_{g_1\sim\rho_1}R(g_1)\leq{}&\mathbb{E}_{g_2\sim\rho_2}r(g_2)-\mathbb{E}_{g_1\sim\rho_1}r(g_1)\\&+a(\lambda)\mathbb{E}_{g_2\sim\rho_2}\mathbb{E}_{g_1\sim\rho_1}\frac{1}{n}\sum_{i=1}^n[\ell(Y_i,g_1(X_i))-\ell(Y_i,g_2(X_i))]^2\\&+b(\lambda)\Bigg[K(\rho_2,\pi_{-\lambda_2 r})+K(\rho_1,\pi_{-\lambda_1 r})+2\log(6\varepsilon^{-1})\\&\quad+\log\mathbb{E}_{g_1\sim\pi_{-\lambda_2 r}}e^{\frac{4\lambda_2^2}{n}\mathbb{E}_{g_2\sim\rho_2}\frac{1}{n}\sum_{i=1}^n[\ell(Y_i,g_1(X_i))-\ell(Y_i,g_2(X_i))]^2}\\&\quad+\log\mathbb{E}_{g_1\sim\pi_{-\lambda_1 r}}e^{\frac{4\lambda_1^2}{n}\mathbb{E}_{g_2\sim\rho_1}\frac{1}{n}\sum_{i=1}^n[\ell(Y_i,g_1(X_i))-\ell(Y_i,g_2(X_i))]^2}\Bigg].\end{aligned}\tag{2.3.2}$$

with $a(\lambda)=\frac{\lambda}{n}\Psi(\frac{\lambda}{n})\big(1+\frac{\lambda}{2n}\big)$ and $b(\lambda)=\frac{2}{\lambda}\Big[1+\frac{\lambda}{n}\Psi(\frac{\lambda}{n})\big(1+\frac{\lambda}{2n}\big)^2\Big]$.

2.3.2. FROM THE EMPIRICAL RELATIVE BOUND TO THE ESTIMATOR. In view of Section 2.2.2, it is natural to concentrate our effort on Gibbs estimators of the form $\pi_{-\lambda r}$ for $\lambda>0$. Introduce for any $0\leq j\leq\log n$ and $\varepsilon>0$,

$$\lambda_j=0.19\sqrt{n}e^{\frac{j}{2}}$$

$$\mathcal{C}(j) = \log \mathbb{E}_{g_1 \sim \pi_{-\lambda_j r}} e^{\frac{4\lambda_j^2}{n} \mathbb{E}_{g_2 \sim \pi_{-\lambda_j r}} \frac{1}{n} \sum_{i=1}^n [\ell(Y_i, g_1(X_i)) - \ell(Y_i, g_2(X_i))]^2}$$
$$L = \log[3 \log^2(en)\varepsilon^{-1}]$$

and for any $0 \leq i < j \leq \log n$ and $\varepsilon > 0$,

$$S(i,j) = a(\lambda_j) \mathbb{E}_{g_1 \sim \pi_{-\lambda_i r}} \mathbb{E}_{g_2 \sim \pi_{-\lambda_j r}} \frac{1}{n} \sum_{i=1}^n [\ell(Y_i, g_1(X_i)) - \ell(Y_i, g_2(X_i))]^2 + b(\lambda_j)\big[\mathcal{C}(i) + \mathcal{C}(j) + 2L\big].$$

Inequality (2.3.2) implies that with probability at least $1 - \varepsilon$, for any $0 \leq i < j \leq \log n$, we have

$$\mathbb{E}_{g_2 \sim \pi_{-\lambda_j r}} R(g_2) - \mathbb{E}_{g_1 \sim \pi_{-\lambda_i r}} R(g_1) \leq \mathbb{E}_{g_2 \sim \pi_{-\lambda_j r}} r(g_2) - \mathbb{E}_{g_1 \sim \pi_{-\lambda_i r}} r(g_1) + S(i,j).$$

This leads me to consider in the chapter 2 of my PhD thesis the following choice of the temperature/complexity parameter in the classification setting.

Algorithm 1. *Let $u(0) = 0$. For any $k \geq 1$, define $\hat{\lambda}_{k-1} = \lambda_{u(k-1)}$ and $u(k)$ as the smallest integer $j \in]u(k-1); \log n]$ such that*

$$\mathbb{E}_{g_2 \sim \pi_{-\lambda_j r}} r(g_2) - \mathbb{E}_{g_1 \sim \pi_{-\hat{\lambda}_{k-1} r}} r(g_1) + S\big(u(k-1), j\big) \leq 0.$$

Classify using a function drawn according to the posterior distribution associated with the last $u(k)$.

This algorithm can be viewed in the following way: it "ranks" the estimator in the model by increasing complexity (if we consider that $K(\pi_{-\lambda_j r}, \pi)$ is the complexity of the estimator associated with $\pi_{-\lambda_j r}$), picks the "first" function in this list and takes at each step the function of smallest complexity such that its risk is smaller than the one at the previous step. This is possible since we have *empirical relative* bounds. Subsequently to this work, different iterative schemes based on empirical relative PAC-Bayesian bounds have been proposed [4, 5, 41]. The interest of the procedure lies in the following theoretical guarantee.

THEOREM 1 *The iterative scheme is finite: there exists $K \in \mathbb{N}$ such that $u(K)$ exists but not $u(K+1)$. With probability at least $1 - \varepsilon$, for any $k \in \{1, \ldots, K\}$, we have*

$$\mathbb{E}_{g \sim \pi_{-\hat{\lambda}_k r}} R(g) \leq \mathbb{E}_{g \sim \pi_{-\hat{\lambda}_{k-1} r}} R(g),$$

and

$$\mathbb{E}_{g \sim \pi_{-\hat{\lambda}_K r}} R(g) \leq \min_{1 \leq j \leq \log n} \left\{ \mathbb{E}_{g \sim \pi_{-\lambda_{j-1} R}} R(g) + C \frac{\log[\log(en)\varepsilon^{-1}]}{\lambda_j} + \frac{1}{\lambda_j} \sup_{0 \leq i \leq j} \left\{ \log \mathbb{E}_{g_1 \sim \pi_{-\lambda_i R}} \mathbb{E}_{g_2 \sim \pi_{-\lambda_i R}} e^{\frac{C\lambda_i^2}{n} \mathbb{P}[g_1(X) \neq g_2(X)]} \right\} \right\}.$$

To illustrate this last theoretical guarantee, let us consider complexity and margin assumptions similar to the ones used in the pioneering work of Mammen and Tsybakov [97]. To detail these assumptions, let d be the (pseudo-)distance on $\mathcal{G}(\mathcal{X};\mathcal{Y})$ defined by

$$d(g_1, g_2) = \mathbb{P}[g_1(X) \neq g_2(X)].$$

Let $\mathcal{G} \subset \mathcal{G}(\mathcal{X};\mathcal{Y})$. For $u > 0$, the set $\mathcal{N} \subset \mathcal{G}(\mathcal{X};\mathcal{Y})$ is called a u-covering net of $\mathcal{G}$ if we have $\mathcal{G} = \cup_{g\in\mathcal{N}}\big\{g' \in \mathcal{G}; d(g, g') \leq u\big\}$. Let $H(u)$ denote the u-covering entropy, i.e. the logarithm of the smallest u-covering net of $\mathcal{G}$. The complexity assumption is that there exist $C' > 0$ and $q > 0$ such that $H(u) \leq C'u^{-q}$ for any $u > 0$. Let

$$g^* = \text{argmin}_{g\in\mathcal{G}} R(g).$$

Without great loss of generality, we assume the existence of such a function. The margin assumption is that there exist $c'', C'' > 0$ and $\kappa \in [1, +\infty]$ such that for any function $g \in \mathcal{G}$,

$$c''\big[R(g) - R(g^*)\big]^{\frac{1}{\kappa}} \leq \mathbb{P}[g(X) \neq g^*(X)] \leq C''\big[R(g) - R(g^*)\big]^{\frac{1}{\kappa}}. \qquad (2.3.3)$$

For any $k \in \mathbb{N}^*$, introduce π_k the uniform distribution on the smallest 2^{-k} covering net.

THEOREM 2 *For the prior distribution $\pi = \sum_{k\geq 1} \frac{\pi_k}{k(k+1)}$, the randomized estimator defined in Algorithm 1 (p.14) satisfies*

$$\mathbb{E}_{g\sim\pi_{-\hat{\lambda}_K r}} R(g) - R(g^*) \leq Cn^{-\frac{\kappa}{2\kappa-1+q}},$$

for some positive constant C.

We also proved in [7, Chap.3, Theorem 3.3] that the right-hand side is the minimax optimal convergence rates under such assumptions. Since the algorithm does not require the knowledge of the margin parameter κ, it is adaptive to this parameter.

Note that Assumption (2.3.3) is stronger than the usual assumption as the latter does not assume the left inequality. In fact, to achieve minimax optimal rates under the usual margin assumption, while still assuming polynomial covering entropies requires the chaining argument [7, Chap.3]. This leads us to study how to combine the chaining argument with the PAC-Bayesian approach and make the connexion with majorizing measures from the generic chaining argument developed by Fernique and Talagrand [120], which we detail in the next section.

2.4. COMBINING PAC-BAYESIAN AND GENERIC CHAINING BOUNDS

There exist many different risk bounds in statistical learning theory. Each of these bounds contains an improvement over the others for certain situations or

algorithms. In [10], Olivier Bousquet and I underline the links between these bounds, and combine several different improvements into a single bound. In particular, we combine the PAC-Bayes approach with the optimal union bound provided by the generic chaining technique developed by Fernique and Talagrand, in a way that also takes into account the variance of the combined functions. We also show how this connects to Rademacher based bounds. The interest in generic chaining rather than just Dudley's chaining [55] comes from the fact that it captures better the behaviour supremum of a Gaussian process [120]. In statistical learning theory, the process of interest and which is asymptotically Gaussian is $g \mapsto R(g) - r(g)$.

I hereafter give a simplified version of the main results of [10]. Let me first introduce the notation. We still consider a set $\mathcal{G} \subset \mathcal{G}(\mathcal{X}; \mathcal{Y})$, $g^* = \text{argmin}_{g \in \mathcal{G}} R(g)$, and that losses take their values in $[0, 1]$. We consider a sequence of nested partitions $(\mathcal{A}_j)_{j \in \mathbb{N}}$ of the set $\mathcal{G}$, that is (i) $\mathcal{A}_j$ is a partition of $\mathcal{G}$ either countable or equal to the set of all singletons of $\mathcal{G}$, and (ii) the $\mathcal{A}_j$ are nested: each element of $\mathcal{A}_{j+1}$ is contained in an element of $\mathcal{A}_j$, and $\mathcal{A}_0 = \{\mathcal{G}\}$. For the partition $\mathcal{A}_j$ and for $g \in \mathcal{G}$, we denote by $A_j(g)$ the unique element of $\mathcal{A}_j$ containing g. Given a sequence of nested partitions $(\mathcal{A}_j)_{j \in \mathbb{N}}$, we can build a collection $(S_j)_{j \in \mathbb{N}}$ of approximating subsets of $\mathcal{G}$ in the following way: for each $j \in \mathbb{N}$, for each element A of $\mathcal{A}_j$, choose a unique element of $\mathcal{G}$ contained in A and define S_j as the set of all chosen elements. We have $|S_0| = 1$ and denote by $p_j(g)$ the unique element of S_j contained in $A_j(g)$. Finally, we also consider that for each $j \in \mathbb{N}$, we have a distribution $\pi^{(j)}$ on $\mathcal{G}$ at our disposal.

Our bound will depend on the specific choices of the distributions $\pi^{(j)}$, the nested partitions $(\mathcal{A}_j)$, the associated sequence of approximating sets (S_j), and the corresponding approximating functions $p_j(g), g \in \mathcal{G}$. Denote δ_g the Dirac measure on g. For a probability distribution ρ on $\mathcal{G}$, define its j-th projection as

$$[\rho]_j = \sum_{g \in S_j} \rho[A_j(g)] \delta_g,$$

when S_j is countable and $[\rho]_j = \rho$ otherwise. For any $\varepsilon > 0$ and $\rho \in \mathcal{M}$, define the complexity of ρ at scale j by

$$\mathcal{K}_j(\rho) = K([\rho]_j, [\pi^{(j)}]_j) + \log[j(j+1)\varepsilon^{-1}],$$

and introduce the average distance between the $(j-1)$-th and j-th projections by

$$\mathcal{D}_j(\rho) = \mathbb{E}_{g \sim \rho} \Bigg\{ \frac{1}{2} \mathbb{E}_{Z \sim P} \Big\{ \ell\big(Y, [p_j(g)](X)\big) - \ell\big(Y, [p_{j-1}(g)](X)\big) \Big\}^2 + \frac{1}{2n} \sum_{i=1}^{n} \Big\{ \ell\big(Y_i, [p_j(g)](X_i)\big) - \ell\big(Y_i, [p_{j-1}(g)](X_i)\big) \Big\}^2 \Bigg\}$$

THEOREM 3 *If the following condition holds*

$$\lim_{j\to+\infty} \sup_{g\in\mathcal{G}} \left\{R(g) - R[p_j(f)] - r(g) + r[p_j(f)]\right\} = 0, \qquad \textit{a.s.} \qquad (2.4.1)$$

then for any $0 < \beta \leq 0.7$*, with probability at least* $1-\varepsilon$*, for any* $\rho \in \mathcal{M}$*, we have*

$$\mathbb{E}_{g\sim\rho} R(g) - R(g^*) \leq \mathbb{E}_{g\sim\rho} r(g) - r(g^*) + \frac{4}{\sqrt{n}} \sum_{j=1}^{+\infty} \sqrt{\mathcal{D}_j(\rho)\mathcal{K}_j(\rho)} + \frac{4}{\sqrt{n}} \sum_{j=1}^{+\infty} \sqrt{\frac{\mathcal{D}_j(\rho)}{\mathcal{K}_j(\rho)}} \log\log\left(4e^2 \frac{\mathcal{K}_j(\rho)}{\mathcal{D}_j(\rho)}\right). \qquad (2.4.2)$$

Assumption (2.4.1) is not very restrictive. For instance, it is satisfied when one of the following condition holds:

- there exists $J \in \mathbb{N}^*$ such that $S_J = \mathcal{G}$,
- almost surely $\lim_{j\to+\infty} \sup_{g\in\mathcal{G}, x\in\mathcal{X}, y\in\mathcal{Y}} |\ell\big(y, g(x)\big) - \ell\big(y, [p_j(g)](x)\big)| = 0$ (it is in particular the case when the bracketing entropy of the set $\mathcal{G}$ is finite for any radius and when the S_j's and p_j's are appropriately built on the bracketing nets of radius going to 0 when $j \to +\infty$).

The bound (2.4.2) combines several previous improvements. It contains an optimal union bound, both in the sense of optimally taking into account the metric structure of the set of functions (via the majorizing measure approach) and in the sense of taking into account the averaging distribution. It is sensitive to the variance of the functions and consequently will lead to fast convergence rates (that is faster than $1/\sqrt{n}$), under margin assumptions such as the ones considered in the works of Ndlec and Massart [100] or Mammen and Tsybakov [97]. It holds for randomized classifiers but contrarily to usual PAC-Bayesian bounds, it remains finite when the averaging distribution is concentrated at a single prediction function. On the negative side, there still remains work in order to get a fully empirical bound (it is not the case here since $\mathcal{D}_j(\rho)$ is not observable) and to better understand the connection with Rademacher averages.

Independently of the generic chaining argument, we use a carefully weighted union bound argument, which is at the origin of the $\log\log$ term in (2.4.2) and leads to the following corollary of the main result in [10].

THEOREM 4 *For any* $\varepsilon > 0$*, with probability at least* $1-\varepsilon$*, for any* $\rho \in \mathcal{M}$*, we have*

$$\mathbb{E}_{g\sim\rho} R(g) - \mathbb{E}_{g\sim\rho} r(g) \leq C\sqrt{\frac{K(\rho,\pi) + \log(2\varepsilon^{-1})}{n}},$$

for some numerical constant $C > 0$ [10, Section 4.3].

This result means that neither the $\log(n)$ term in (McA) (p.8) or the Shannon's entropy term in (2.1.3) (p.6) is needed if we are allowed to have a numerical factor slightly larger in front of the square root term.

Chapter 3

The three aggregation problems

3.1. INTRODUCTION

Aggregation is about combining different prediction functions in order to get a better prediction. It has become popular and has been intensively studied these last two decades partly thanks to the success of boosting algorithms, and principally of the AdaBoost algorithm, introduced by Freund and Schapire [58]. These algorithms use linear combination of a large number of simple functions to provide a classification decision rule.

In this chapter, we focus on the least squares setting, in which the outputs are real numbers and the risk of a prediction function $g : \mathcal{X} \to \mathbb{R}$ is

$$R(g) = \mathbb{E}[Y - g(X)]^2.$$

Our results are nevertheless of interest for classification also as any estimate of the conditional expectation of the output knowing the input leads by thresholding to a classification decision rule, and the quality of this plug-in estimator is directly linked to the quality of the least squares regression estimator (see [53, Section 6.2], [16] and specifically the comparison lemmas of its section 5, and also [95, 27, 28] for consistency results in classification using other surrogate loss functions).

Boosting type classification methods usually aggregate simple functions, but the aggregation is also of interest when some potentially complicated functions are aggregated. More precisely, when facing the data, the statistician has often to choose several models which are likely to be relevant for the task. These models can be of similar structures (like embedded balls of functional spaces) or on the contrary of very different nature (e.g., based on kernels, splines, wavelets or on parametric approaches). For each of these models, we assume that we have a learning scheme which produces a 'good' prediction function in the sense that its risk is as small as the risk of the best function of the model up to some small additive term[1]. Then the question is to decide on how we use or combine/aggregate these schemes. One possible answer is to split the data into two groups, use the first group to train the prediction function (i.e. compute the estimator) associated with each model, and then use the second group to build a prediction function which is as good as (i) the best of the previously learnt prediction functions, (ii) the best convex combination of these functions or (iii) the best linear combination of these functions, in terms of risk, up to some small additive term. The three aggregation problems we will focus on in this chapter concern the second part

[1]The learning procedure could differ for each model, or on the contrary, be the same but using different values of a tuning parameter.

of this scheme. The idea of mixing (or combining or aggregating) the estimators originally appears in [110, 71, 132, 133].

We hereafter treat the initial estimators as fixed functions, which means that the results hold conditionally on the data set on which they have been obtained, this data set being independent of the n input-output observations Z_1^n. Specifically, let $g_1,\ldots,g_d$ be d prediction functions, with $d \geq 2$. Introduce

$$g^*_{\mathbf{MS}} \in \operatorname*{argmin}_{g\in\{g_1,\ldots,g_d\}} R(g),$$

$$g^*_{\mathbf{C}} \in \operatorname*{argmin}_{g\in\{\sum_{j=1}^d \theta_j g_j;\theta_1\geq 0,\ldots,\theta_d\geq 0,\sum_{j=1}^d \theta_j=1\}} R(g),$$

and

$$g^*_{\mathbf{L}} \in \operatorname*{argmin}_{g\in\{\sum_{j=1}^d \theta_j g_j;\theta_1\in\mathbb{R},\ldots,\theta_d\in\mathbb{R}\}} R(g).$$

The model selection aggregation task **(MS)** is to find an estimator $\hat{g}$ based on the observed data Z_1^n for which the excess risk $R(\hat{g}) - R(g^*_{\mathbf{MS}})$ is guaranteed to be small. Similarly, the convex (resp. linear) aggregation task **(C)** (resp. **(L)**) is to find an estimator $\hat{g}$ for which the excess risk $R(\hat{g}) - R(g^*_{\mathbf{C}})$ (resp. $R(\hat{g}) - R(g^*_{\mathbf{L}})$) is guaranteed to be small.

The minimax optimal rates of aggregation are given in [123] and references within. Under suitable assumptions, it is shown that there exist estimators $\hat{g}_{\mathbf{MS}}$, $\hat{g}_{\mathbf{C}}$ and $\hat{g}_{\mathbf{L}}$ such that

$$\mathbb{E}R(\hat{g}_{\mathbf{MS}}) - R(g^*_{\mathbf{MS}}) \leq C\min\left(\frac{\log d}{n}, 1\right), \tag{3.1.1}$$

$$\mathbb{E}R(\hat{g}_{\mathbf{C}}) - R(g^*_{\mathbf{C}}) \leq C\min\left(\sqrt{\frac{\log(1+d/\sqrt{n})}{n}}, \frac{d}{n}, 1\right),$$

$$\mathbb{E}R(\hat{g}_{\mathbf{L}}) - R(g^*_{\mathbf{L}}) \leq C\min\left(\frac{d}{n}, 1\right),$$

where $\hat{g}_L$ (and for $d \leq n$, $\hat{g}_C$) require the knowledge of the input distribution. We recall that C is a positive constant that may differ from line to line. Tsybakov [123] has shown that these rates cannot be uniformly improved in the following sense. Let $\sigma > 0$, $L > 0$ and Let $\mathcal{P}_{\sigma,L}$ be the set of probability distributions on $\mathcal{X}\times\mathbb{R}$ such that we almost surely have $Y = g(X) + \xi$, with $\|g\|_\infty \leq L$, and ξ a centered Gaussian random variable independent of X and with variance σ^2. For appropriate choices of $g_1,\ldots,g_d$, the following lower bounds hold:

$$\inf_{\hat{g}} \sup_{P\in\mathcal{P}_{\sigma,L}} \left\{\mathbb{E}R(\hat{g}) - R(g^*_{\mathbf{MS}})\right\} \geq C\min\left(\frac{\log d}{n}, 1\right),$$

$$\inf_{\hat{g}} \sup_{P \in \mathcal{P}_{\sigma,L}} \left\{ \mathbb{E}R(\hat{g}) - R(g^*_{\mathbf{C}}) \right\} \geq C \min \left(\sqrt{\frac{\log(1+d/\sqrt{n})}{n}}, \frac{d}{n}, 1 \right),$$

$$\inf_{\hat{g}} \sup_{P \in \mathcal{P}_{\sigma,L}} \left\{ \mathbb{E}R(\hat{g}) - R(g^*_{\mathbf{L}}) \right\} \geq C \min \left(\frac{d}{n}, 1 \right),$$

where the infimum is taken over all estimators. The three aggregation tasks have also been studied in the least squares regression with fixed design, where similar rates are obtained [36, 50, 51].

This chapter will provide my contributions to the aggregation problems (in the random design setting) summarized as follows.

- The expected excess risk $\mathbb{E}R(\hat{g}) - R(g^*_{\mathbf{MS}})$ of the empirical risk minimizer on $\{g_1, \ldots, g_d\}$ (or its penalized variants) cannot be uniformly smaller than $C\sqrt{\frac{\log d}{n}}$. Since the minimax optimal rate is $\frac{\log d}{n}$, this shows that these estimators are inappropriate for the model selection task (Section 3.2.1).

- Catoni [39] and Yang [131] have independently shown that the optimal rate $\frac{\log d}{n}$ in the model selection problem is achieved for the progressive mixture rule. In [9], I provide a variant of this estimator coming from the field of sequential prediction of nonrandom sequences, and called the progressive indirect mixture rule. It has the benefit of satisfying a tighter excess risk bound in a bounded setting (outputs in $[-1, 1]$). I also study the case when the outputs have heavy tails (much thicker than exponential tails), and show how the noise influences the minimax optimal convergence rate. I also provide refined lower bounds of Assouad's type with tight constants (Section 3.2.2).

- In [8], I show a limitation of the algorithms known to satisfy (3.1.1): despite having an expected excess risk of order $1/n$ (if we drop the dependence in d), the excess risk of the progressive (indirect or not) mixture rule suffers deviations of order $1/\sqrt{n}$ (Section 3.2.3).

- This last result leads me to define a new estimator $\hat{g}$ which does not suffer from this drawback: the deviations of the excess risk $\mathbb{E}R(\hat{g}) - R(g^*_{\mathbf{MS}})$ is of order $\frac{\log d}{n}$ (Section 3.2.4).

- In my PhD (its first chapter), I provide an estimator $\hat{g}$ based on empirical bounds of any aggregation procedures for which with high probability

$$R(\hat{g}) - R(g^*_{\mathbf{C}}) \leq \begin{cases} C\sqrt{\frac{\log(d \log n)}{n}} & \text{always,} \\ C\frac{\log(d \log n)}{n} & \text{if } R(g^*_{\mathbf{MS}}) = R(g^*_{\mathbf{C}}). \end{cases}$$

This means that for $n^{\frac{1}{2}+\delta} \leq d \leq e^n$ with $\delta > 0$, the estimator has the minimax optimal rate of task **(C)**, and is adaptive to the extent that it has also the minimax optimal rate of task **(MS)** when $R(g^*_{\mathbf{MS}}) = R(g^*_{\mathbf{C}})$ (Section 3.3).

- Finally, Olivier Catoni and I [14] provide minimax results for **(L)**, and consequently also for **(C)** when $d \leq \sqrt{n}$. The strong point of these results is that it does not require the knowledge of the input distribution, nor uniformly bounded exponential moments of the conditional distribution of the output knowing the input and has no extra logarithmic factor unlike previous results. In particular, provided that we know H and σ such that $\|g^*_{\mathbf{L}}\|_\infty \leq H$ and $\sup_{x\in\mathcal{X}} \mathbb{E}\big\{[Y - g^*_{\mathbf{L}}(X)]^2 \big| X = x\big\} \leq \sigma^2$, we propose an estimator $\hat{g}$ satisfying $\mathbb{E}R(\hat{g}) - R(g^*_{\mathbf{L}}) \leq 68(\sigma + H)^2 \, \frac{d+2}{n}$ (Section 3.4).

I should conclude this introductory section by emphasizing that we will not assume that the regression function $g^{(\text{reg})} : x \mapsto \mathbb{E}(Y|X = x)$, which minimizes the risk functional, is in the linear span of $\{g_1, \ldots, g_d\}$. This means that bounds of the form

$$\mathbb{E}R(\hat{g}) - R(g^*) \leq c[R(g^{(\text{reg})}) - R(g^*)] + \text{residual term}, \tag{3.1.2}$$

with $c > 1$ are not of interest in our setting[2], as they would not provide the minimax learning rate when $R(g^{(\text{reg})}) \gg R(g^*)$.

3.2. Model selection type aggregation

3.2.1. Suboptimality of empirical risk minimization. Any empirical risk minimizer and any of its penalized variants are really poor algorithms in this task since their expected convergence rate cannot be uniformly faster than $\sqrt{(\log d)/n}$. The following lower bound comes from [8] (see [92], [39, p.14], [90, 72, 106] for similar results and variants).

Theorem 5 *For any training set size n, there exist d prediction functions $g_1, \ldots, g_d$ taking their values in $[-1, 1]$ such that for any learning algorithm $\hat{g}$ producing a prediction function in $\{g_1, \ldots, g_d\}$ there exists a probability distribution generating the data for which $Y \in [-1, 1]$ almost surely, and*

$$\mathbb{E}R(\hat{g}) - R(g^*_{\mathbf{MS}}) \geq \min\Big(\sqrt{\frac{\lfloor \log_2 d \rfloor}{4n}}, 1\Big),$$

[2]These last bounds, which are relatively common in the literature, are nonetheless useful in a nonparametric setting in which the statistician is allowed to take $\{g_1, \ldots, g_d\}$ large enough so that $R(g^{(\text{reg})}) - R(g^*)$ is of the same order as the residual term.

where $\lfloor \log_2 d \rfloor$ denotes the largest integer smaller or equal to the logarithm in base 2 of d.

3.2.2. PROGRESSIVE INDIRECT MIXTURE RULES. The result of the previous section shows that, to obtain the minimax optimal rate given in (3.1.1), an estimator has to look at an enlarged set of prediction functions. Until our work, the only known optimal estimator was based on a Cesaro mean of Bayesian estimators, also referred to as progressive mixture rule.

To define it, let π be the uniform distribution on the finite set $\{g_1, \ldots, g_d\}$. For any $i \in \{0, \ldots, n\}$, the cumulative loss suffered by the prediction function g on the first i pairs of input-output, denoted Z_1^i for short, is

$$\Sigma_i(g) = \sum_{k=1}^{i} [Y_k - g(X_k)]^2,$$

where by convention we take Σ_0 identically equal to zero. Let $\lambda > 0$ be a parameter of the estimator. Recall that $\pi_{-\lambda\Sigma_i}$ is the distribution on $\{g_1, \ldots, g_d\}$ admitting a density with respect to π that is proportional to $e^{-\lambda\Sigma_i}$.

The *progressive mixture rule* (PM) predicts according to $\frac{1}{n+1}\sum_{i=0}^{n} \mathbb{E}_{g\sim\pi_{-\lambda\Sigma_i}}\, g$. In other words, for a new input x, the predicted output is

$$\frac{1}{n+1}\sum_{i=0}^{n} \frac{\sum_{j=1}^{d} g_j(x) e^{-\lambda\Sigma_i(g_j)}}{\sum_{j=1}^{d} e^{-\lambda\Sigma_i(g_j)}}.$$

A specificity of PM is that its proof of optimality is not achieved by the most prominent tool in statistical learning theory: bounds on the supremum of empirical processes (see [125], and refined works as [26, 80, 99, 34] and references within). The idea of the proof, which comes back to Barron [24], is based on a chain rule and appeared to be successful for least squares and entropy losses [38, 39, 25, 131] and for general loss in [72].

Here my first contribution was to take ideas coming from the field of sequential prediction of nonrandom sequences (see e.g. [107, 46] for a general overview and [65, 44, 45, 134] for more specific results with sharp constants) and propose a slight generalization of progressive mixture rules, that I called progressive indirect mixture rules.

The *progressive indirect mixture rule* (PIM) is also parameterized by a real number $\lambda > 0$, and is defined as follows. For any $i \in \{0, \ldots, n\}$, let $\hat{h}_i$ be a prediction function such that

$$[Y - \hat{h}_i(X)]^2 \leq -\frac{1}{\lambda} \log \mathbb{E}_{g\sim\pi_{-\lambda\Sigma_i}}\, e^{-\lambda[Y-g(X)]^2} \qquad \text{a.s.} \qquad (3.2.1)$$

If one of the $\hat{h}_i$ does not exist, the algorithm is said to fail. Otherwise it predicts according to $\frac{1}{n+1}\sum_{i=0}^{n}\hat{h}_i$.

This estimator is a direct transposition from the sequential prediction algorithm proposed and studied in [126, 65, 127] to our "batch" setting. The functions $\hat{h}_i$ do not necessarily exist, but are also not necessarily unique when they exist. A technical justification of (3.2.1) comes from the analysis of PM synthetically written in Appendix D.

When $\max\big(|Y|, |g_1(X)|, \ldots, |g_d(X)|\big) \leq B$ a.s. for some $B > 0$ and for λ large enough, the functions $\hat{h}_i$ exist (so the algorithm does not fail). Still in this uniformly bounded setting, it can be shown that PM is a PIM for λ large enough. On the other hand, there exists $\lambda > 0$ small enough for which the algorithm does not fail and such that PM is not a particular case of PIM, that is one cannot take $\hat{h}_i = \mathbb{E}_{g\sim\pi_{-\lambda\Sigma_i}}\, g$ to satisfy (3.2.1) (see [65, Example 3.13]). In fact, it is also shown there that PIM will not generally produce a prediction function in the convex hull of $\{g_1, \ldots, g_d\}$ unlike PM. The following amazingly sharp upper bound on the excess risk of PIM holds.

THEOREM 6 *Assume that $|Y| \leq 1$ a.s. and $\|g_j\|_\infty \leq 1$ for any $j \in \{1, \ldots, d\}$. Then, for $\lambda \leq \frac{1}{2}$,* PIM *does not fail and its expected excess risk is upper bounded by $\frac{\log d}{\lambda(n+1)}$, that is*

$$\mathbb{E}_{Z_1^n} R\left(\frac{1}{n+1}\sum_{i=0}^{n}\hat{h}_i\right) - R(g^*_{\mathbf{MS}}) \leq \frac{\log d}{\lambda(n+1)}. \tag{3.2.2}$$

It essentially comes from a result in sequential prediction and the fact that results expressed in cumulative loss can be transposed to our setting since the expected risk of the randomized procedure based on sequential predictions is proportional to the expectation of the cumulative loss of the sequential procedure. Precisely, the following statement holds.

LEMMA 7 *Let $\mathcal{A}$ be a learning algorithm which produces the prediction function $\mathcal{A}(Z_1^i)$ at time $i+1$, i.e. from the data $Z_1^i = (Z_1, \ldots, Z_i)$. Let $\mathcal{L}$ be the randomized algorithm which produces a prediction function $\mathcal{L}(Z_1^n)$ drawn according to the uniform distribution on $\{\mathcal{A}(\emptyset), \mathcal{A}(Z_1), \ldots, \mathcal{A}(Z_1^n)\}$. The (doubly) expected risk of $\mathcal{L}$ is equal to $\frac{1}{n+1}$ times the expectation of the cumulative loss of $\mathcal{A}$ on the sequence $Z_1, \ldots, Z_{n+1}$, where Z_{n+1} denotes a random variable independent of the training set $Z_1^n = (Z_1, \ldots, Z_n)$ and with the same distribution P.*

My second contribution to model selection aggregation in [9] is to provide a different viewpoint of the progressive mixture rule from the one in [72], leading to a slight improvement in the moment condition of the initial version of [72]. The

result is the following and is extended to the L_q loss functions for $q \geq 1$ in [9, Section 7].

THEOREM 8 *Assume that $\|g_j\|_\infty \leq 1$ for any $j \in \{1, \ldots, d\}$, and $\mathbb{E}|Y|^s \leq A$ for some $s \geq 2$ and $A > 0$. For $\lambda = C_1\big(\frac{\log d}{n}\big)^{2/(s+2)}$ with $C_1 > 0$, the expected excess risk of PM is upper bounded by $C\big(\frac{\log d}{n}\big)^{s/(s+2)}$, that is*

$$\mathbb{E}_{Z_1^n} R\left(\frac{1}{n+1}\sum_{i=0}^{n} \mathbb{E}_{g\sim\pi_{-\lambda\Sigma_i}}\, g\right) - R(g^*_{\mathbf{MS}}) \leq C\left(\frac{\log d}{n}\right)^{s/(s+2)},$$

for a quantity C which depends only on C_1, A and s.

The convergence rate cannot be improved in a minimax sense [9, Section 8.3.2]. These results show how heavy output tails influence the learning rate: for the limiting case $s = 2$, the bounds are of order $n^{-1/2}$ while for s going to infinity, it is of order of n^{-1}, that is the rate in the bounded case, or in the uniformly bounded conditional exponential moment setting.

The lower bounds developed to prove the minimax optimality of the above result are based on a refinement of Assouad's lemma, which allows to get much tighter constants. For instance, they improve the lower bounds for Vapnik-Cervonenkis classes [53, Chapter 14] by a factor greater than 1000, and lead to the following simple bound.

THEOREM 9 *Let $\mathcal{F}$ be a set of binary classification functions of VC-dimension V. For any classification rule $\hat{f}$ trained on a data set of size $n \geq \frac{V}{4}$, there exists a probability distribution generating the data for which*

$$\mathbb{E}R(\hat{f}) - \inf_{f\in\mathcal{F}} R(f) \geq \frac{1}{8}\sqrt{\frac{V}{n}}. \tag{3.2.3}$$

3.2.3. LIMITATION OF PROGRESSIVE INDIRECT MIXTURE RULES. Let $\hat{g}_\lambda$ denote a progressive indirect mixture rule (it could be a progressive mixture or not) for some $\lambda > 0$. Under boundedness assumptions (and even under some exponential moment assumptions) and appropriate choice of λ, $\hat{g}_\lambda$ satisfies an expected excess risk bound of order $\frac{\log d}{n}$. Then one would also expect the excess risk $R(\hat{g}) - R(g^*_{\mathbf{MS}})$ to be of order $\frac{\log d}{n}$ with high probability. In fact, this does not necessarily happen as the following theorem holds for $d = 2$.

THEOREM 10 *Let g_1 and g_2 be the constant functions respectively equal to 1 and -1. For any $\lambda > 0$ and any training set size n large enough, there exist $\varepsilon > 0$ and*

a distribution generating the data for which $Y \in [-1,1]$ almost surely, and with probability larger than ε, we have

$$R(\hat{g}_\lambda) - R(g^*_{\mathbf{MS}}) \geq c\sqrt{\frac{\log(e\varepsilon^{-1})}{n}}$$

where c is a positive constant only depending on λ.

More precisely, in [8], it is shown that for large enough n, and some constants $c_1 > 1/2$ and $c_2 > 0$ only depending on λ, with probability at least $1/n^{c_1}$, we have $R(\hat{g}_\lambda) - R(g^*_{\mathbf{MS}}) \geq c_2\sqrt{(\log n)/n}$. Since $c_1 > 1/2$, there is naturally no contradiction with the fact that, in expectation, the excess risk is of order $\frac{\log d}{n}$.

3.2.4. GETTING ROUND THE PREVIOUS LIMITATION. I now present the algorithm introduced in [8], and called the empirical star estimator, which has both expectation and deviation convergence rate of order $\frac{\log d}{n}$. The empirical risk of a prediction function $g : \mathcal{X} \to \mathbb{R}$ is defined by

$$r(g) = \frac{1}{n}\sum_{i=1}^{n}[Y_i - g(X_i)]^2.$$

Let $\hat{g}^{(\text{erm})}$ be an empirical risk minimizer among the reference functions:

$$\hat{g}^{(\text{erm})} \in \operatorname*{argmin}_{g\in\{g_1,\ldots,g_d\}} r(g).$$

For any prediction functions g, g', let $[g, g']$ denote the set of functions which are convex combination of g and g': $[g, g'] = \big\{\alpha g + (1-\alpha)g' : \alpha \in [0,1]\big\}$. The empirical star estimator $\hat{g}^{(\text{star})}$ minimizes the empirical risk over a star-shaped set of functions, precisely:

$$\hat{g}^{(\text{star})} \in \operatorname*{argmin}_{g\in[\hat{g}^{(\text{erm})},g_1]\cup\cdots\cup[\hat{g}^{(\text{erm})},g_d]} r(g).$$

The main result concerning this estimator is the following.

THEOREM 11 *Assume that $|Y| \leq B$ almost surely and $\|g_j\|_\infty \leq B$ for any $j \in \{1,\ldots,d\}$. Then the empirical star algorithm satisfies: for any $\varepsilon > 0$, with probability at least $1-\varepsilon$,*

$$R(\hat{g}^{(\text{star})}) - R(g^*_{\mathbf{MS}}) \leq \frac{200B^2\log[3d(d-1)\varepsilon^{-1}]}{n} \leq \frac{600B^2\log(d\varepsilon^{-1})}{n}.$$

Consequently, we also have

$$\mathbb{E}R(\hat{g}^{(\text{star})}) - R(g^*_{\mathbf{MS}}) \leq \frac{400B^2\log(3d)}{n}.$$

An additional advantage of this empirical star estimator is that it does not need to know the constant B. In other words, it is adaptive to the smallest value of B for which the boundedness assumptions hold. This was not the case of the progressive mixture rules in which we need to take $\lambda \leq 1/(2B^2)$ for the indirect ones and $\lambda \leq 1/(8B^2)$ for the "direct" one in order to state Inequality (3.2.2). On the negative side, the theoretical guarantee on the expected excess risk is 200 times larger than the one stated for the best PIM. However, this is more an artefact of the intricate proof of Theorem 11 than a drawback of the algorithm.

Another difference between progressive mixture rules is that the function output by the estimator is inside $\cup_{1\leq j<k\leq d}[g_j, g_d]$, which is not in general the case for the progressive (indirect) mixture rules. We have already seen in Section 3.2.1 that the empirical risk minimizer on $\{g_1, \ldots, g_d\}$ has not the minimax optimal rate. A natural question in view of the empirical star algorithm is whether empirical risk minimization on $\cup_{1\leq j<k\leq d}[g_j, g_d]$ would reach the $(\log d)/n$ rate. It can be proved for $d = 3$ that, even under boundedness assumptions, the rate cannot be better than $n^{-2/3}$ for an adequate choice of the functions and the distribution (proof omitted by lack of interest in negative results).

Interestingly, Lecu and Mendelson [91] proposed a variant of the empirical star algorithm, which also uses the empirical risk minimizer $\hat{g}^{(\text{erm})}$ to define a set of functions on which the empirical risk is minimized. Precisely, for a confidence level $\varepsilon > 0$, let $\hat{\mathcal{G}}$ be the set of functions $g \in \{g_1, \ldots, g_d\}$ satisfying

$$r(g) \leq r(\hat{g}^{(\text{erm})}) + CB\sqrt{\frac{\log(2d\varepsilon^{-1})}{n}}\sqrt{\frac{\sum_{i=1}^n [g(X_i) - \hat{g}^{(\text{erm})}(X_i)]^2}{n}} + C\frac{B^2 \log(2d\varepsilon^{-1})}{n}.$$

where C is a positive constant. The final estimator is the empirical risk minimizer in the convex hull of $\hat{\mathcal{G}}$. It is also shown there that the selection of a subset of functions $\hat{\mathcal{G}}$ before taking the convex hull is necessary to achieve the minimax convergence rate since the empirical risk minimizer on the convex hull of $\{g_1, \ldots, g_d\}$ has an excess risk at least of order $1/\sqrt{n}$ for an appropriate distribution and d of order $\sqrt{n}$.

The advantage of the empirical star algorithm over the empirical risk minimizer on the convex hull of $\hat{\mathcal{G}}$ is its adaptivity to both the confidence level and the constant B, and a theoretical guarantee of the form $C\frac{\log(d\varepsilon^{-1})}{n}$ instead of $C\frac{\log(d)\log(\varepsilon^{-1})}{n}$ for the empirical risk minimizer on the convex hull of $\hat{\mathcal{G}}$.

When $d \leq \sqrt{n}$, the minimax learning rate for problem **(C)** and **(L)** are both of order $\frac{d}{n}$, meaning that estimators solving problem **(L)** are solutions to problem **(C)** for $d \leq \sqrt{n}$. So, estimators for $d \leq \sqrt{n}$ are given in the section devoted to linear aggregation (Section 3.4), and this section focuses on the case when $d \geq n^{\frac{1}{2}+\delta}$.

The literature contains few results for problem **(C)** with constant $c = 1$ in (3.1.2) and minimax optimal residual term for $d \geq n^{\frac{1}{2}+\delta}$, with $\delta > 0$. The first type of results is to apply the progressive mixture rule on an appropriate grid of the simplex [123]. Another solution is to use the exponentiated gradient algorithm introduced and studied by Kivinen and Warmuth [74] in the context of sequential prediction for the quadratic loss, and then extended to general loss functions by Cesa-Bianchi [43]. Lemma 7 has to be invoked to convert these algorithms and the bounds to our statistical framework. Juditsky, Nazin, Tsybakov and Vayatis [73] has viewed the resulting algorithm as a stochastic version of the mirror descent algorithm, and proposed a different choice of the temperature parameter, while still reaching the optimal convergence rate. All the above results hold in expectation, and it is not clear that the deviations of the excess risk bounds are sub-exponential. The estimator presented hereafter does not share this drawback.

To address problem **(C)** (defined in page 20), the first chapter of my PhD thesis establishes empirical excess risk bounds for any estimator that produces a prediction function in the convex hull of $g_1, \ldots, g_d$ whatever the empirical data are. Any such estimator $\hat{g}$ can be associated with a function $\hat{\rho}$ mapping a training set to a distribution on $\{g_1, \ldots, g_d\}$ such that $\hat{g}(Z_1^n) = \mathbb{E}_{g\sim\hat{\rho}(Z_1^n)} g$. Conversely, any mapping $\hat{\rho}$ from $\mathcal{Z}^n$ (the set of training sets of size n) to the set $\mathcal{M}$ of distributions on $\{g_1, \ldots, g_d\}$ defines an estimator

$$\hat{g} = \mathbb{E}_{g\sim\hat{\rho}} g,$$

where we have dropped the training set Z_1^n for sake of compactness. Similarly, there exists a distribution $\rho_{\mathbf{C}}^*$ on $\{g_1, \ldots, g_d\}$ such that

$$g_{\mathbf{C}}^* = \mathbb{E}_{g\sim\rho_{\mathbf{C}}^*} g.$$

The assumptions are boundedness of the functions $g_1, \ldots, g_d$ and of the regression function $g^{(\text{reg})} : x \mapsto \mathbb{E}(Y|X = x)$ and uniform boundedness of the conditional exponential moments of the output knowing the input. Precisely, there exist $B > 0$, $\alpha > 0$, and $M > 0$ such that for any g', g'' in $\{g^{(\text{reg})}, g_1, \ldots, g_d\}$, $\|g' - g''\|_\infty \leq B$ and for any $x \in \mathcal{X}$,

$$\mathbb{E}\big(e^{\alpha|Y-g^{(\text{reg})}(X)|}\big|X = x\big) \leq M.$$

THEOREM 12 *Under the above assumptions, there exist $C_1, C_2 > 0$ depending only on the constant M and the product αB such that for any (prior) distribution $\pi \in \mathcal{M}$, any $\varepsilon > 0$, and any aggregating procedure $\hat{\rho} : \mathcal{Z}^n \to \mathcal{M}$, with probability at least $1 - \epsilon$,*

$$R(\mathbb{E}_{g\sim\hat{\rho}} g) - R(g^*_{\mathbf{C}}) \leq \min_{\lambda\in[0,C_1]} \Bigg\{ (1+\lambda)\big[r(\mathbb{E}_{g\sim\hat{\rho}} g) - r(g^*_{\mathbf{C}})\big] + \frac{2\lambda}{n}\sum_{i=1}^{n} \mathbb{V}\mathrm{ar}_{g\sim\hat{\rho}}\, g(X_i) + C_2\frac{B^2}{n}\frac{K(\hat{\rho},\pi) + \log(2\log(2n)\varepsilon^{-1})}{\lambda} \Bigg\}. \tag{3.3.1}$$

This bound comes from the PAC-Bayesian analysis, and consequently, the complexity of an aggregating procedure is measured by the Kullback-Leibler divergence of $\hat{\rho}$ with respect to some prior distribution π on $\{g_1, \ldots, g_d\}$. In absence of prior knowledge, π is chosen as the uniform distribution, which allows to bound uniformly the KL divergence by $\log d$. Besides the usual empirical excess risk, Inequality (3.3.1) depends on the empirical variance of $g(x)$ when g is drawn according to $\hat{\rho}$. Unlike the Kullback-Leibler term, this term is small for concentrated posterior distributions.

All previous results of this chapter were easily generalizable to loss of quadratic type under boundedness assumptions, that is loss with second derivative with respect to its second argument uniformly lower and upper bounded by positive constants. To my knowledge, the generalization cannot be done here as the analysis strongly relies on the remarkable identity[3]

$$R(\mathbb{E}_{g\sim\rho} g) = \mathbb{E}_{(g',g'')\sim\rho\otimes\rho}\mathbb{E}[Y - g'(X)][Y - g''(X)], \tag{3.3.2}$$

which is specific to the quadratic loss and allows to apply the PAC-Bayesian analysis for distributions on the product space $\{g_1, \ldots, g_d\} \times \{g_1, \ldots, g_d\}$.

Let $\hat{\rho}_{\mathbf{C}}$ be the distribution minimizing the right-hand side of (3.3.1) with π the uniform distribution on $\{g_1, \ldots, g_d\}$ and where $-(1+\lambda)r(g^*_{\mathbf{C}})$ is replaced by its upper bound $-r(g^*_{\mathbf{C}}) - \lambda \min_{g\in\{\sum_{j=1}^d \theta_j g_j;\theta_1\geq 0,\ldots,\theta_d\geq 0,\sum_{j=1}^d \theta_j = 1\}} r(g)$. When defining $\hat{\rho}_{\mathbf{C}}$, for sake of computability of the estimator [7, Chap.1, Theorem 4.2.2], one can also replace the minimum over $[0, C_1]$ by a minimum over a geometric grid of the interval $[n^{-1}, C_1]$ without altering the validity of the following theorem.

THEOREM 13 *For any $\varepsilon > 0$, with probability at least $1 - \epsilon$, we have*

$$R(\mathbb{E}_{g\sim\hat{\rho}_{\mathbf{C}}} g) - R(g^*_{\mathbf{C}}) \leq CB\sqrt{\frac{\log(d\log(2n)\varepsilon^{-1})}{n}\mathbb{E}\,\mathbb{V}\mathrm{ar}_{g\sim\rho^*_{\mathbf{C}}}\, g(X)}$$

[3] To be precise, [7, Chap.1] used $R(\mathbb{E}_{g\sim\rho} g) = \mathbb{E}_{g\sim\rho} R(g) - \frac{1}{2}\mathbb{E}_{g'\sim\rho}\mathbb{E}_{g''\sim\rho}\mathbb{E}[g'(X) - g''(X)]^2$, but it would have been more direct to use (3.3.2).

$$+ CB^2 \frac{\log(d \log(2n) \varepsilon^{-1})}{n},$$

for some constant $C > 0$ depending only on αB and M.

Since $\mathbb{E}\, \mathbb{V}\mathrm{ar}_{g \sim \rho_{\mathbf{C}}^*} g(X) \leq B^2/4$, the excess risk is at most of order $\sqrt{\frac{\log(d \log(2n))}{n}}$, that is the minimax convergence rate of the convex aggregation task for $d \geq n^{\frac{1}{2}+\delta}$, with $\delta > 0$. Besides, when the best convex combination occurs to be a vertex of the simplex defined by $\{g_1, \ldots, g_d\}$ the variance term equals zero, and thus, the convergence rate is $\frac{\log(d \log(2n))}{n}$, that is the minimax convergence rate of model selection type aggregation (at least for $d \geq \log(2n)$).

3.4. LINEAR AGGREGATION

To handle problems **(C)** and **(L)** in the same framework and also to incorporate other possible constraints on the coefficients of the linear combination, let us consider Θ a closed convex subset of $\mathbb{R}^d$, and define

$$\mathcal{G} = \left\{ \sum_{j=1}^{d} \theta_j g_j ; (\theta_1, \ldots, \theta_d) \in \Theta \right\}.$$

Introduce the vector-valued function $\vec{g} : x \mapsto (g_1(x), \ldots, g_d(x))^T$. The function $\sum_{j=1}^d \theta_j g_j$ can then be simply written $\langle \theta, \vec{g} \rangle$ with $\theta = (\theta_1, \ldots, \theta_d)^T$. Let

$$g^* \in \operatorname*{argmin}_{g \in \mathcal{G}} R(g).$$

Thus, when Θ is the simplex of $\mathbb{R}^d$, we have $g^* = g_{\mathbf{C}}^*$ and when $\Theta = \mathbb{R}^d$, we have $g^* = g_{\mathbf{L}}^*$.

Aggregating linearly functions to design a prediction function with low quadratic risk is just the problem of linear least squares regression. It is a central task in statistics, since both linear parametric models and nonparametric estimation with linear approximation spaces (piecewise polynomials based on a regular partition, wavelet expansions, trigonometric polynomials, ...) are popular. It has thus been widely studied.

Classical statistical textbooks often only state results for the fixed design setting as a bound of order d/n can be rather easily obtained in this case. This can be misleading since it does not imply a d/n upper bound in the random design setting. For the truncated ordinary least squares estimator, Györfi, Kohler, Krzyżak and Walk [63, Theorem 11.3] give a bound of the form of (3.1.2, page 22) with

residual term of order $\frac{d \log n}{n}$ and $c = 8$. When the input distribution is known, Tsybabov [123] provides a bound of order d/n on the expected risk of a projection estimator on an orthonormal basis of $\mathcal{G}$ for the dot product $(f, g) \mapsto \mathbb{E}[f(X)g(X)]$.

Catoni [37, Proposition 5.9.1] and Alquier [5] have used the PAC-Bayesian approach to prove high probability excess risk bounds of order d/n involving the conditioning of the Gram matrix $Q = \mathbb{E}\left[\vec{g}(X)\vec{g}(X)^T\right]$. Both results require at least exponential moments on the conditional distribution of the output Y knowing the input vector $\vec{g}(X)$.

It can be derived from the work of Birgé and Massart [31] an excess risk bound for the empirical risk minimizer of order at worst $\frac{d \log n}{n}$, and asymptotically of order d/n. It holds with high probability, for a bounded set Θ and requires bounded input vectors and conditional exponential moments of the output. Localized Rademacher complexities [80, 26] also allows to study the empirical risk minimizer on a bounded set of functions. They lead to a high probability d/n convergence rate of the empirical risk minimizer under strong assumptions: uniform boundedness of the input vector, the output and the parameter set Θ.

Penalized least squares estimators using the L^2-norm of the vector of coefficients, or more recently, its L^1-norm have also been widely studied. A common characteristic of the excess risk bounds obtained for these estimators is that it is of order d/n only under strong assumptions on the eigenvalues (of submatrices) of Q.

In [14], Olivier Catoni and I provide new risk bounds for the ridge estimator and the ordinary least squares estimator (Section 3.4.1). We also propose a min-max estimator which has non-asymptotic guarantee of order d/n under weak assumptions on the distributions of the output Y and the random variables $g_j(X)$, $j = 1, \dots, d$ (Section 3.4.2). Finally, we propose a sophisticated PAC-Bayesian estimator which satisfies a simpler d/n bound (Section 3.4.3).

The key common surprising factor of these results is the absence of exponential moment condition on the output distribution while achieving exponential deviations. All risk bounds are obtained through a PAC-Bayesian analysis on truncated differences of losses. Our results tend to say that truncation leads to more robust algorithms. Local robustness to contamination is usually invoked to advocate the removal of outliers, claiming that estimators should be made insensitive to small amounts of spurious data. Our work leads to a different theoretical explanation. The observed points having unusually large outputs when compared with the (empirical) variance should be down-weighted in the estimation of the mean, since they contain less information than noise. In short, huge outputs should be truncated because of their low signal to noise ratio.

3.4.1. RIDGE REGRESSION AND EMPIRICAL RISK MINIMIZATION. The ridge regression estimator on $\mathcal{G}$ is defined by $\hat{g}^{(\text{ridge})} = \langle \hat{\theta}^{(\text{ridge})}, \vec{g} \rangle$ with

$$\hat{\theta}^{(\text{ridge})} \in \arg\min_{\theta \in \Theta} r(\langle \theta, \vec{g} \rangle) + \lambda \|\theta\|^2,$$

where λ is some nonnegative real parameter and $r(\langle \theta, \vec{g} \rangle)$ is the empirical risk of the function $\langle \theta, \vec{g} \rangle$. In the case when $\lambda = 0$, the ridge regression $\hat{g}^{(\text{ridge})}$ is nothing but the empirical risk minimizer $\hat{g}^{(\text{erm})}$.

In the same way we consider the optimal ridge function optimizing the expected ridge risk: $\tilde{g} = \langle \tilde{\theta}, \vec{g} \rangle$ with

$$\tilde{\theta} \in \arg\min_{\theta \in \Theta} \big\{ R(\langle \theta, \vec{g} \rangle) + \lambda \|\theta\|^2 \big\}.$$

Our first result is of asymptotic nature. It is stated under weak hypotheses, taking advantage of the weak law of large numbers.

THEOREM 14 *Let us assume that*

$$\mathbb{E}\big[\|\vec{g}(X)\|^4\big] < +\infty, \tag{3.4.1}$$

$$\textit{and} \quad \mathbb{E}\Big\{ \|\vec{g}(X)\|^2 \big[\tilde{g}(X) - Y\big]^2 \Big\} < +\infty. \tag{3.4.2}$$

Let $\nu_1, \ldots, \nu_d$ be the eigenvalues of the Gram matrix $Q = \mathbb{E}\big[\vec{g}(X)\vec{g}(X)^T\big]$, and let $Q_\lambda = Q + \lambda I$ be the ridge regularization of Q. Let us define the effective ridge dimension

$$D = \sum_{i=1}^{d} \frac{\nu_i}{\nu_i + \lambda} \mathbb{1}_{\nu_i > 0} = \operatorname{Tr}\big[(Q + \lambda I)^{-1} Q\big] = \mathbb{E}\big[\|Q_\lambda^{-1/2} \vec{g}(X)\|^2\big].$$

When $\lambda = 0$, D is equal to the rank of Q and is otherwise smaller. For any $\varepsilon > 0$, there is n_ε, such that for any $n \geq n_\varepsilon$, with probability at least $1 - \varepsilon$,

$$R(\hat{g}^{(\text{ridge})}) + \lambda \|\hat{\theta}^{(\text{ridge})}\|^2 \leq \min_{\theta \in \Theta} \big\{ R(\langle \theta, \vec{g} \rangle) + \lambda \|\theta\|^2 \big\} + C \operatorname{ess\,sup} \mathbb{E}\big\{ [Y - \tilde{g}(X)]^2 \big| X \big\} \frac{D + \log(3\varepsilon^{-1})}{n},$$

for some numerical constant $C > 0$.

This theorem shows that the ordinary least squares estimator (obtained when $\Theta = \mathbb{R}^d$ and $\lambda = 0$), as well as the empirical risk minimizer on any closed convex set, asymptotically reach a d/n speed of convergence under very weak hypotheses. It shows also the regularization effect of the ridge regression. There

emerges an *effective dimension* D, where the ridge penalty has a threshold effect on the eigenvalues of the Gram matrix.

On the other hand, the weakness of this result is its asymptotic nature : n_ε may be arbitrarily large under such weak hypotheses, and this shows even in the simplest case of the estimation of the mean of a real-valued random variable by its empirical mean, which is the case when $d = 1$ and $\vec{g}(X) \equiv 1$ [42]. Typically, the proof of Theorem 14 shows that n_ε is of order $1/\varepsilon$. To avoid this limitation, we were conducted to consider more involved algorithms as described in the following two sections.

3.4.2. A MIN-MAX ESTIMATOR FOR ROBUST ESTIMATION. This section provides an alternative to the empirical risk minimizer with non asymptotic exponential risk deviations of order d/n for any confidence level. Moreover, we will assume only a second order moment condition on the output and cover the case of unbounded inputs, the requirement on the random variables $g_j(X)$ being only a finite fourth order moment. On the other hand, we assume that the set Θ of the vectors of coefficients is bounded. (This still allows to solve problem **(L)** as soon as we know a bounded set in which $g^*_{\mathbf{L}}$ lies for sure.)

Let $\alpha > 0$ and consider the truncation function:

$$T(x) = \begin{cases} -\log\bigl(1 - x + x^2/2\bigr) & 0 \le x \le 1, \\ \log(2) & x \ge 1, \\ -T(-x) & x \le 0, \end{cases}$$

For any $g, g' \in \mathcal{G}$, introduce

$$\mathcal{D}(g, g') = \sum_{i=1}^{n} T\Bigl(\alpha\bigl[Y_i - g(X_i)\bigr]^2 - \alpha\bigl[Y_i - g'(X_i)\bigr]^2\Bigr).$$

Let us assume in this section that for any $j \in \{1, \dots, d\}$,

$$\mathbb{E}\bigl\{g_j(X)^2[Y - g^*(X)]^2\bigr\} < +\infty, \tag{3.4.3}$$

and

$$\mathbb{E}\bigl[g_j^4(X)\bigr] < +\infty. \tag{3.4.4}$$

Define

$$\mathcal{S} = \{g \in \operatorname{span}\{g_1, \dots, g_d\} : \mathbb{E}[g(X)^2] = 1\}, \tag{3.4.5}$$

$$\sigma = \sqrt{\mathbb{E}\bigl\{[Y - g^*(X)]^2\bigr\}} = \sqrt{R(g^*)}, \tag{3.4.6}$$

$$\chi = \max_{g \in \mathcal{S}} \sqrt{\mathbb{E}[g(X)^4]}, \tag{3.4.7}$$

$$\kappa = \frac{\sqrt{\mathbb{E}\big\{[\vec{g}(X)^T Q^{-1}\vec{g}(X)]^2\big\}}}{\mathbb{E}\big[\vec{g}(X)^T Q^{-1}\vec{g}(X)\big]}, \tag{3.4.8}$$

$$\kappa' = \frac{\sqrt{\mathbb{E}\big\{[Y-g^*(X)]^4\big\}}}{\mathbb{E}\big\{[Y-g^*(X)]^2\big\}} = \frac{\sqrt{\mathbb{E}\big\{[Y-g^*(X)]^4\big\}}}{\sigma^2}, \tag{3.4.9}$$

$$\mathcal{R} = \max_{g',g''\in\mathcal{G}} \sqrt{\mathbb{E}\big\{[g'(X)-g''(X)]^2\big\}}. \tag{3.4.10}$$

THEOREM 15 *Let us assume that* (3.4.3) *and* (3.4.4) *hold. For some numerical constants* c *and* c'*, for*

$$n > c\kappa\chi d,$$

by taking

$$\alpha = \frac{1}{2\chi\big[2\sqrt{\kappa'}\sigma + \sqrt{\chi}\mathcal{R}\big]^2}\left(1 - \frac{c\kappa\chi d}{n}\right), \tag{3.4.11}$$

for any estimator $\hat{g}$ *satisfying* $\hat{g} \in \mathcal{G}$ *a.s., for any* $\varepsilon > 0$*, with probability at least* $1-\varepsilon$*, we have*

$$\begin{aligned} R(\hat{g}) - R(g^*) \leq & \frac{1}{n\alpha}\left(\max_{g'\in\mathcal{G}} \mathcal{D}(\hat{g},g') - \inf_{g\in\mathcal{G}} \max_{g'\in\mathcal{G}} \mathcal{D}(g,g')\right) \\ & + \frac{c\kappa\kappa' d\sigma^2}{n} + \frac{8\chi\big(\frac{\log(\epsilon^{-1})}{n} + \frac{c'\kappa^2 d^2}{n^2}\big)\big[2\sqrt{\kappa'}\sigma + \sqrt{\chi}\mathcal{R}\big]^2}{1 - \frac{c\kappa\chi d}{n}}. \end{aligned}$$

The above theorem suggest to look for function realizing the min-max of $(g,g') \mapsto \mathcal{D}(g,g')$. More precisely, an estimator such that

$$\max_{g'\in\mathcal{G}} \mathcal{D}(\hat{g},g') < \inf_{g\in\mathcal{G}} \max_{g'\in\mathcal{G}} \mathcal{D}(g,g') + \sigma^2\frac{d}{n},$$

has a non asymptotic bound for the excess risk with a d/n convergence rate and an exponential tail even when neither the output Y nor the input vector $\vec{g}(X)$ has exponential moments. This stronger non asymptotic bound compared to the bounds of the previous section comes at the price of replacing the empirical risk minimizer by a more involved estimator. Nevertheless, reasonable heuristics can be developed to compute it approximately [14, Section 3], and leads to a significantly better estimator of $g^*_{\mathbf{L}}$ than the ordinary least squares estimator when there is some heavy-tailed noise (see Appendix G).

3.4.3. A SIMPLE TIGHT RISK BOUND FOR A SOPHISTICATED PAC-BAYES ALGORITHM. A disadvantage of the min-max estimator proposed in the previous

section is that its theoretical guarantee depends (implicitly) on kurtosis like coefficients. We provide in [14, Section 4] a more sophisticated estimator, having the following simple excess risk bound independent of these kurtosis like quantities, and still of order $\frac{d}{n}$. It holds under stronger assumption on the input vector $\vec{g}(X)$ (precisely, uniform boundedness), still assumes that the set Θ is bounded, and holds under a second order moment condition on the output.

THEOREM 16 *Assume that $\mathcal{G}$ has a diameter H for L^∞-norm:*

$$\sup_{g',g''\in\mathcal{G},x\in\mathcal{X}} |g'(x)-g''(x)| = H \tag{3.4.12}$$

and that, for some $\sigma > 0$,

$$\sup_{x\in\mathcal{X}} \mathbb{E}\big\{[Y-g^*(X)]^2\big|X=x\big\} \le \sigma^2 < +\infty.$$

There exists an estimator $\hat{g}$ such that for any $\varepsilon > 0$, with probability at least $1-\varepsilon$, we have

$$R(\hat{g}) - R(g^*) \le 17(2\sigma+H)^2\,\frac{d+\log(2\varepsilon^{-1})}{n}.$$

On the negative side, when the target is to solve problem **(L)**, it requires the knowledge of a L^∞-bounded ball in which f^*_{lin} lies and an upper bound on $\sup_{x\in\mathcal{X}} \mathbb{E}\big\{[Y-f^*_{\text{lin}}(X)]^2\big|X=x\big\}$. The looser this knowledge is, the bigger the constant in front of d/n is. On the positive side, the convergence rate is of order d/n, without neither extra logarithmic factor, nor constant factors involving the conditioning of the Gram matrix Q or some Kurtosis like coefficients.

To conclude this section, let us add that, when the output admits uniformly bounded conditional exponential moments, a relatively simple Gibbs estimator also achieves the d/n convergence rate. Precisely we have the following theorem.

THEOREM 17 *Assume that* (3.4.12) *holds for $H < +\infty$, and that there exist $\alpha > 0$ and $M > 0$ such that for any $x \in \mathcal{X}$,*

$$\mathbb{E}\big(e^{\alpha|Y-g^*_{\mathbf{L}}(X)|}\big|X=x\big) \le M.$$

Consider the probability distribution $\hat{\pi}$ on $\mathcal{G}$ defined by its density with respect to the uniform distribution π on $\mathcal{G}$:

$$\frac{\hat{\pi}}{\pi}(g) = \frac{e^{-\lambda\sum_{i=1}^n [Y_i-g(X_i)]^2}}{\mathbb{E}_{g'\sim\pi} e^{-\lambda\sum_{i=1}^n [Y_i-g'(X_i)]^2}},$$

where $\lambda > 0$ is appropriately chosen (depending on α, H and M). For any $\varepsilon > 0$, with probability at least $1-\varepsilon$, we have

$$R(\mathbb{E}_{g\sim\hat{\pi}} g) - R(g^*) \le C\,\frac{d+\log(2\varepsilon^{-1})}{n},$$

where the quantity $C > 0$ only depends on α, H and M.

From the minimax rates of the three aggregation problems, we see that for $n \ll d \ll e^n$, one can predict as well as the best convex combination up to a small additive term, which is at most of order $\sqrt{\frac{\log d}{n}}$, but one cannot expect to predict in general as well as the best linear combination up to a small additive term. In this setting, one may want to reduce its target by trying to predict as well as (still up to a small additive term) the best linear combination of at most $s \ll d$ functions, that is the function

$$g^* \in \operatorname*{argmin}_{g \in \{\sum_{j=1}^d \theta_j g_j ; \theta_1 \in \mathbb{R}, \dots, \theta_d \in \mathbb{R}, \sum_{j=1}^d \mathbb{1}_{\theta_j \neq 0} \leq s\}} R(g). \tag{3.5.1}$$

It is well-established that L^1 regularization allows to perform this task. The procedure is known as Lasso [122, 113] and is defined by $\hat{f}^{(\text{lasso})} = \langle \hat{\theta}^{(\text{lasso})}, \vec{g} \rangle$ with

$$\hat{\theta}^{(\text{lasso})} \in \operatorname*{argmin}_{\theta \in \mathbb{R}^d} \frac{1}{n} \sum_{i=1}^n \big(Y_i - \langle \theta, \vec{g}(X_i) \rangle \big)^2 + \lambda \|\theta\|_1,$$

where $\lambda > 0$ is a parameter to be tuned to retrieve the desired number of relevant variables/functions[4]. As the L^2 penalty used in ridge regression, the L^1 penalty shrinks the coefficients. The difference is that for coefficients which tend to be close to zero, the shrinkage makes them equal to zero. This allows to select relevant variables/functions (i.e., find the j's such that $\theta_j^* \neq 0$).

If we assume that the regression function $g^{(\text{reg})}$ is a linear combination of only $s \ll d$ variables among $\{g_1, \dots, g_d\}$, the typical result is to prove that the expected excess risk of the Lasso estimator for λ of order $\sqrt{(\log d)/n}$ is of order $(s \log d)/n$ [36, 124, 105, 93]. Since this quantity is much smaller than d/n, this makes a huge improvement (provided that the sparsity assumption is true). This kind of results usually requires strong conditions on the eigenvalues of submatrices of Q, essentially assuming that the functions g_j are near orthogonal. Here we will argue that by combining the estimators solving **(MS)** and **(L)**, one can achieve minimax optimal learning rate without requiring such conditions. The guarantees presented here are also stronger than the ones associated with L_0-regularization (penalization proportional to the number of nonzero coefficient) whatever criterion (Mallows' C_p [96], AIC [3] or BIC [116]) is used to tune the penalty constant. Recent advances on theoretical guarantees of L_0-regularization can be found in the works of Bunea, Tsybakov and Wegkamp [36] and of Birg and Massart [32] for

[4] The functions $g_1, \dots, g_d$ can be called the explanatory variables of the output. Note also that we can consider without loss of generality that the input space is $\mathbb{R}^d$ and that the functions $g_1, \dots, g_d$ are the coordinate functions.

the fixed design setting and in the work of Raskutti, Wainwright and Yu [114] for the random design setting considered here. These results for L_0-regularization are not as good for the ones for the estimator described in this section since the $(s \log d)/n$ excess risk bound holds only when the conditional expectation of the output knowing the input is inside the model.

Precisely, let us assume[5] that for some $B > 0$, $\|g^*\|_\infty \leq B$ and $|Y| \leq B$. Let $\mathcal{L}_1$ denote the first half of the training set $\{Z_1, \ldots, Z_{n/2}\}$, and $\mathcal{L}_2$ denote the second half of the training set $\{Z_{n/2+1}, \ldots, Z_n\}$, where for simplicity we have assumed that n is even. For any $I \subset \{1, \ldots, d\}$ of size s, let $\hat{g}_I$ be the sophisticated estimator that satisfies Theorem 16 *trained on* $\mathcal{L}_1$ and associated with the set $\mathcal{G}_I = \big\{\langle \theta, \vec{g} \rangle : \|\langle \theta, \vec{g} \rangle\|_\infty \leq B, \theta_j = 0 \text{ for any } j \notin I\big\}$. (One can alternatively consider the Gibbs estimator of Theorem 17.) Let $\hat{g}$ be the empirical star estimator (defined in Section 3.2.4) *trained on* $\mathcal{L}_2$ and associated with the $\binom{d}{s}$ functions $\hat{g}_I$ (that are non-random given $\mathcal{L}_1$). This two-stage estimator satisfies the following theorem.

THEOREM 18 *For any $\varepsilon > 0$, with probability at least $1 - \varepsilon$,*

$$R(\hat{g}) - R(g^*) \leq CB^2 \frac{s \log(d/s) + \log(2\varepsilon^{-1})}{n}, \tag{3.5.2}$$

for some numerical constant $C > 0$.

PROOF. From Theorem 11, since we have $\binom{d}{s} \leq \big(ed/s\big)^s$, with probability at least $1 - \varepsilon/2$, we have

$$R(\hat{g}) - \min_{I \subset \{1,\ldots,d\} : |I| = s} R(\hat{g}_I) \leq 1200 B^2 \frac{s \log(ed/s) + \log(2\varepsilon^{-1})}{n}.$$

Let I^* be a set of s variables containing the set of at most s variables involved in g^*. From Theorem 16, with probability at least $1 - \varepsilon/2$, we have

$$R(\hat{g}_{I^*}) - R(g^*) \leq 1224 B^2 \frac{s + \log(4\varepsilon^{-1})}{n}.$$

By using an union bound, we obtain

$$R(\hat{g}) - R(g^*) \leq 1224 B^2 \left(\frac{s \log(ed/s) + \log(2\varepsilon^{-1})}{n} + \frac{s + \log(4\varepsilon^{-1})}{n} \right).$$

which gives the desired result. □

[5] We make boundedness assumptions for sake of simplicity. The results can be generalized to outputs having exponential conditional moments since both building blocks of the estimator can handle this type of noisy outputs: for the empirical star algorithm, see the supplemental material of [8]. Further generalizations are open problems.

Due to the particular structure of the empirical star algorithm, the estimator $\hat{g}$ can be written as a linear combination of at most $2s$ functions among $\{g_1, \ldots, g_d\}$, so that the estimator can be used for variable selection. The functions involved in g^* do not necessarily belong to this set of at most $2s$ functions. I do not believe that achieving such identifiability of these particular relevant variables should be the goal, since pursuing this target would definitely require that the different variables are not too much correlated, a situation which will rarely occurs in practice.

Adaptivity with respect to the sparsity level of g^* can also be obtained. Indeed, let s^* be the number of nonzero coefficients of the function g^* defined by (3.5.1). By using a three-stage estimator procedure using successively the empirical star algorithm at a given level of sparsity and then on the s functions thus designed, it is easy that (3.5.2) still holds with s replaced by s^*. Note that g^* defined in (3.5.1) depends on a sparsity level s, which can be taken equal to d. Then we have $g^* = g^*_{\mathbf{L}}$, and the three-stage procedure is adaptive to the sparsity level of $g^*_{\mathbf{L}}$. In the fixed design setting, Bunea, Tsybakov and Wegkamp [36] have shown that these rates are minimax optimal, and it is natural to consider that their lower bound extends to our random design case.

Another possible use of the algorithms solving problems **(MS)** and **(L)** is when we consider sparsity with group structure. This occurs when the variables are naturally organized into groups: in computer vision, this naturally occurs since there exist different families of image descriptors, and the grouping can be done by family, scale and/or position. Let $I_1, \ldots, I_D \subset \{1, \ldots, d\}$ be D sets of grouped variables. For a vector θ, let us say that a group I_k is active if there exists $j \in I_k$ such that $\theta_j \neq 0$. let $S(\theta)$ be the number of active groups among $I_1, \ldots, I_D$.

For a given sparsity level $s \in \{1, \ldots, D\}$, the target is

$$g^{(\text{group})} \in \operatorname*{argmin}_{g \in \{\langle \theta, \vec{g} \rangle ; \theta \in \mathbb{R}^d, S(\theta) \leq s\}} R(g).$$

There exist only $\binom{D}{s}$ different sets of s groups that could be active. So a two-stage estimator $\hat{g}^{(\text{group})}$ similar to the one described before satisfies that with probability at least $1 - \varepsilon$,

$$R(\hat{g}^{(\text{group})}) - R(g^{(\text{group})}) \leq CB^2 \frac{s \log(D/s) + J + \log(2\varepsilon^{-1})}{n},$$

where J denotes the number of nonzero coefficients in the linear combination defining $g^{(\text{group})}$. This type of results has not been obtained yet for the group Lasso [135] even when assuming low correlation between the variables, except for the fixed design setting [69, 94].

We have presented in this section an example of theoretical results easily obtainable from the estimators solving problems **(MS)** and **(L)**. The results are expressed in terms of sub-exponential excess risk bounds, which were not obtainable

before the introduction of the empirical star algorithm. An advantage of the approach is its genericity: it is not restricted to particular families of estimators.

There are yet some limitations. First, there is no variable selection consistency with this approach, but as stated before, this stronger type of results would require strong assumptions on the input vector distribution, that are often not met in practice. In the fixed design setting, for overlapping groups, Jenatton, Bach and I [70] have proved a high dimensional variable consistency result extending the corresponding result for the Lasso [138, 128].

Second, the approach does not extend easily to the case of generalized additive models, in which linear combinations of a fixed number of functions are replaced by functional spaces [104], such as reproducing kernel Hilbert spaces in the cases of multiple kernel learning [86, 23, 111, 108, 22, 81].

Finally, the most important limitation, which is often encountered when using classical model selection approach, is its computational intractability. So this leaves open the following fundamental problem: is it possible to design a computationally efficient algorithm with the above guarantees (i.e., without assuming low correlation between the explanatory variables)?

Chapter 4

Multi-armed bandit problems

4.1. INTRODUCTION

Bandit problems illustrate the fundamental difficulty of decision making in the face of uncertainty: a decision maker must choose between following what seems to be the best choice in view of the past ("exploiting") or testing ("exploring") some alternative, hoping to discover a choice that beats the current best choice. More precisely, in the multi-armed bandit problem, at each stage, an agent (or decision maker) chooses one action (or arm), and receives a reward from it. The agent aims at maximizing his rewards. Since he does not know the process generating the rewards, he needs to explore (try) the different actions and yet, exploit (concentrate its draws on) the seemingly most rewarding arms.

The multi-armed bandit problem is the simplest setting where one encounters the exploration-exploitation dilemma. It has a wide range of applications including advertizement [21, 52], economics [29, 85], games [59] and optimization [77, 48, 76, 35]. It can be a central building block of larger systems, like in evolutionary programming [68] and reinforcement learning [119], in particular in large state space Markovian Decision Problems [79]. The name "bandit" comes from imagining a gambler in a casino playing with K slot machines, where at each round, the gambler pulls the arm of any of the machines and gets a payoff as a result. The seminal work of Robbins [115] casts the bandit problem in a stochastic setting in which essentially the rewards obtained from an arm are independent and identically distributed random variables that are also independent from the rewards obtained from the other arms. Since the work of Auer, Cesa-Bianchi, Freund and Schapire [19], it was also studied in an adversarial setting.

To set the notation, let $K \geq 2$ be the number of actions (or arms) and $n \geq K$ be the time horizon. A K-armed bandit problem is a game between an agent and an environment in which, at each time step $t \in \{1, \ldots, n\}$, (i) the agent chooses a probability distribution p_t on a finite set $\{1, \ldots, K\}$, (ii) the environment chooses a reward vector $g_t = (g_{1,t}, \ldots, g_{K,t}) \in [0,1]^K$ (possibly through some external randomization), and simultaneously (independently), the agent draws the arm I_t according to the distribution p_t, (iii) the agent only gets to see his own reward $g_{I_t,t}$. The goal of the decision maker is to maximize his cumulative reward $\sum_{t=1}^{n} g_{I_t,t}$.

In the stochastic bandit problem, the environment cannot choose any reward vectors: the reward vectors g_t have to be independent and identically distributed, and its components should be independent random variables[1]. So an environment

[1]The independence of the components is always made in the literature, but is not fundamentally useful (up to rare modifications of the numerical constants).

is just parameterized by a K-tuple of probability distributions $(\nu_1, \ldots, \nu_K)$ on $[0, 1]$. Note that the term "stochastic bandit" can be a bit misleading since the assumption is not just stochasticity but rather an i.i.d. assumption.

In the adversarial bandit problem, no such restriction is put so that past gains have no reason to be representative of future ones. This contrasts with the stochastic setting in which confidence bounds on the mean reward of the arms can be deduced from the rewards obtained so far.

A policy is a strategy for choosing the drawing probability distribution p_t based on the history formed by the past plays and the associated rewards. So it is a function that maps any history to a probability distribution on $\{1, \ldots, K\}$. We define the regret of a policy with respect to the best constant decision as

$$R_n = \max_{i=1,\ldots,K} \sum_{t=1}^{n} \big(g_{i,t} - g_{I_t,t}\big). \tag{4.1.1}$$

To compare to the best constant decision is a reasonable target since it is well-known that (i) there exist randomized policies ensuring that $\mathbb{E}R_n/n$ tends to zero as n goes to infinity, (ii) this convergence property would not hold if the maximum and the sum would be inverted in the definition of R_n. This chapter will first present my contributions to the stochastic bandit problems, essentially:

- how to use empirical variance estimates in upper confidence based policies? (Section 4.2.4)
- how thin is the tail distribution of the regret of standard policies, and how can we improve it? (Section 4.2.5)
- provide a minimax optimal policy (Section 4.2.6),
- propose a model and an arm-increasing rule to deal with bandit problems with more arms than draws: $K \geq n$ (Section 4.2.7),
- design and use a Bernstein's bound with estimated variances to have better stopping rules (Section 4.2.8),
- provide a policy to identify the best arm at the end of the n time steps (Section 4.2.9).

Sbastien Bubeck and I [12] contribute to the adversarial setting by designing a new type of weighted average forecaster characterized by an implicit normalization of the weights, and for which a new type of analysis can be developed. The advantage of the policy and the analysis is that it allows to bridge the long open logarithmic gap in the characterization of the minimax rate for the multi-armed

bandit problem, and to have a common framework for addressing other sequential prediction problems (full information, label efficient, tracking the best expert) (Section 4.3).

4.2. THE STOCHASTIC BANDIT PROBLEM

4.2.1. NOTATION. Let $T_i(t)$ denote the number of times arm i is chosen by the policy during the first t plays. Define $\mu_i = \int x\nu_i(dx)$ the expectation and $V_i = \int (x-\mu_i)^2\nu_i(dx)$ the variance of the distribution ν_i characterizing arm i. Let $i^* \in \text{argmin}_{i\in\{1,\dots,K\}}\mu_i$ denote an index of an optimal arm. The suboptimality of an arm i is measured by:

$$\Delta_i = \max_{j=1,\dots,K} \mu_j - \mu_i = \mu_{i^*} - \mu_i.$$

Let $X_{i,t}$ be the t-th reward obtained from arm i if $T_i(n) \geq t$, and for $t > T_i(n)$, let $X_{i,t}$ be other independent realizations of ν_i. For any $i \in \{1,\dots,K\}$ and $s \in \mathbb{N}$, introduce $\overline{X}_{i,s}$ and $\bar{V}_{i,s}$ the empirical mean and variance of $X_{i,1},\dots,X_{i,s}$.

$$\overline{X}_{i,s} = \frac{1}{s}\sum_{j=1}^{s} X_{i,j} \quad \text{and} \quad \bar{V}_{i,s} = \frac{1}{s}\sum_{j=1}^{s}(X_{i,j} - \overline{X}_{i,s})^2.$$

4.2.2. REGRET NOTION. Previous works in the stochastic bandit problem do not use the regret defined by (4.1.1), which is a regret with respect to the best constant decision, but a (pseudo-)regret that compares the reward of the policy to the reward of an optimal arm in expectation, that is $i^* \in \text{argmin}_{i\in\{1,\dots,K\}}\mu_i$:

$$\overline{R}_n = \sum_{t=1}^{n} \left(g_{i^*,t} - g_{I_t,t}\right) \leq R_n.$$

Results concerning this regret are easier to state, and we will follow hereafter the trend of previous works to state the results in terms of $\overline{R}_n$. In this section, we gather results showing how to go from an upper bound on $\overline{R}_n$ to an upper bound on R_n. The following lemma shows that logarithmic regret bounds on $\mathbb{E}\overline{R}_n$ extend to logarithmic regret bounds on $\mathbb{E}R_n$ *when the optimal arm is unique*, that is $\mu_i < \mu_{i^*}$ for any $i \neq i^*$. Besides, unlike known upper bounds for $\mathbb{E}\overline{R}_n$, the ones on $\mathbb{E}R_n$ depends on the variance V_{i^*} of the reward distribution of the optimal arm. (When there are several optimal arms, it is the smallest variance of the optimal arms distributions which appears in the expected regret bound.)

LEMMA 19 ([12]) *For a given $\delta \geq 0$, let $I = \big\{i \in \{1,\dots,K\} : \Delta_i \leq \delta\big\}$ be the set of arms "δ-close" to the optimal ones, and $J = \{1,\dots,K\} \setminus I$ the remaining set of arms. In the stochastic bandit game, we have*

$$\mathbb{E}R_n - \mathbb{E}\overline{R}_n \leq \sqrt{\frac{n\log|I|}{2}} + \sum_{i\in J} \frac{1}{2\Delta_i}\exp(-n\Delta_i^2),$$

and also

$$\mathbb{E}R_n - \mathbb{E}\overline{R}_n \leq \sqrt{\frac{n\log|I|}{2}} + \sum_{i\in J} \frac{2V_{i^*} + 2V_i + 2\Delta_i/3}{\Delta_i}\exp\left(-\frac{n\Delta_i^2}{2V_{i^*} + 2V_i + 2\Delta_i/3}\right).$$

In particular when there exists a unique arm i^* such that $\Delta_{i^*} = 0$, we have

$$\mathbb{E}R_n - \mathbb{E}\overline{R}_n \leq 2\sum_{i\neq i^*} \frac{V_{i^*} + V_i + \Delta_i/3}{\Delta_i},$$

and also for any $t > 0$

$$\mathbb{P}\big(R_n - \overline{R}_n > t\big) \leq \sum_{i\neq i^*} \exp\left(-\frac{(t+n\Delta_i)^2}{n\min(1, 2V_{i^*} + 2V_i + 2(t/n+\Delta_i)/3)}\right).$$

The uniqueness of the optimal arm is really needed to have logarithmic (in n) bounds on the expected regret. This can be easily seen by considering a two-armed bandit in which both reward distributions are identical (and non degenerated). In this case, the expected pseudo-regret is equal to zero while the expected regret will be at least of order $\sqrt{n}$ for any forecaster. This reveals a fundamental difference between the expected regret and the pseudo-regret.

Previous works on stochastic bandits use the expected pseudo-regret criterion since it satisfies

$$\mathbb{E}\overline{R}_n = \sum_{i=1}^{K} \Delta_i \mathbb{E}T_i(n),$$

meaning that one has only to control the expected sampling times of suboptimal arms to understand how the expected pseudo-regret behaves.

4.2.3. INTRODUCTION TO UPPER CONFIDENCE BOUNDS POLICIES. Early papers have studied stochastic bandit problems under Bayesian assumptions (e.g., Gittins [61]). On the contrary, Lai and Robbins [84] have considered a parametric minimax framework. They have introduced an algorithm that follows what is now called the "optimism in the face of uncertainty principle". At time $t \equiv k_t \pmod K$ with $k_t \in \{1,\dots,K\}$, their policy compares an *upper confidence bound* (UCB)

of the mean reward μ_{k_t} of arm k_t to a reasonable target defined as the highest empirical mean of "sufficiently" drawn arms. If the upper confidence bound exceeds the target, arm k_t is drawn, and otherwise, the arm defining the reasonable target is drawn. Lai and Robbins proved that the expected regret of this policy increases at most at a logarithmic rate with the number of trials and that the algorithm achieves the smallest possible regret up to some sub-logarithmic additive term (for the considered family of distributions). Agrawal [2] proposed computationally easier UCB algorithms in a more general setting that have also logarithmic expected regret (at the price of a higher numerical constant in the upper bound on the regret). More recently, Auer, Cesa-Bianchi and Fischer [18] have proposed even simpler policies achieving logarithmic regret *uniformly over time* rather than just for a fixed number n of rounds known in advance by the agent. Besides, unlike previous works, they have provided non asymptotic bounds.

Upper confidence bounds policies can be described as follows. From time 1 to K, draw each arm once. At time $t \geq K+1$, draw the arm maximizing $B_{i,T_i(t-1),t}$, where $B_{i,s,t}$ is a high probability bound on μ_i computed from the i.i.d. sample $X_{i,1}, \ldots, X_{i,s}$. The confidence level of this high probability bound might depend on the current round t. For instance, the UCB1 policy of Auer, Cesa-Bianchi and Fischer [18] uses

$$B_{i,s,t} = \overline{X}_{i,s} + \sqrt{\frac{2\log t}{s}},$$

which is an upper bound on μ_i holding with probability at least $1 - t^{-4}$ according to Hoeffding's inequality.

Auer, Cesa-Bianchi and Fischer [18] also noted that plugging an upper confidence bound of the variance in the square root term performs empirically substantially better than UCB1. Precisely, their experiments used

$$B_{i,s,t} = \overline{X}_{i,s} + \sqrt{\min\left(\bar{V}_{i,s} + \sqrt{\frac{2\log t}{s}}, \frac{1}{4}\right)\frac{\log t}{s}}. \tag{4.2.1}$$

My first contribution to the multi-armed bandit problem was to provide a theoretical justification of these empirical findings, as described in the following section.

4.2.4. UCB POLICY WITH VARIANCE ESTIMATES. Rmi Munos, Csaba Szepesvári and I [15] have proposed the following slight modification of the arm indexes given by (4.2.1):

$$B_{i,s,t} = \overline{X}_{i,s} + \sqrt{\frac{2\zeta\bar{V}_{i,s}\log t}{s}} + \frac{3\zeta\log t}{s}, \tag{4.2.2}$$

with $\zeta > 1$. The associated policy achieves a logarithmic regret as UCB1 with a constant factor potentially much smaller than the one of UCB1. Indeed, from

[18], UCB1 satisfies

$$\mathbb{E}\overline{R}_n \leq \sum_{i:\Delta_i>0} \frac{10}{\Delta_i} \log n, \tag{4.2.3}$$

whereas our algorithm, called UCB-V (V for variance), satisfies for $\zeta > 1$,

$$\mathbb{E}\overline{R}_n \leq c_\zeta \sum_{i:\Delta_i>0} \left(\frac{\sigma_k^2}{\Delta_k} + 2 \right) \log n, \tag{4.2.4}$$

with $c_\zeta > 0$ a function of ζ satisfying $c_{1.2} \leq 10$ and $c_\zeta \leq C\Big(\sum_{t=1}^{+\infty} t^{-\zeta} + \zeta \Big)$ for some numerical constant $C > 0$. We also proved that for specific distributions of the rewards, UCB-V with $\zeta < 1$ suffers a polynomial expected (pseudo-)regret, that is $\mathbb{E}\overline{R}_n \geq Cn^C$ for some $C > 0$. The argument proving this later assertion also implies that using exactly the upper bound (4.2.1) can dramatically fail in some specific situations[2].

4.2.5. DEVIATION OF THE REGRET OF UCB POLICIES. In this section, we consider that there is a unique optimal arm i^*. In [15], we show that the UCB-V policy defined by (4.2.2) satisfies

$$\mathbb{P}(\overline{R}_n \geq C \log n) \leq \left(\frac{C'}{\log n} \right)^{\zeta/2}. \tag{4.2.5}$$

for quantities C and C' depending on $K, \zeta, \sigma_1, \ldots, \sigma_K, \Delta_1, \ldots, \Delta_K$, but not on n. The "polynomial" rate in (4.2.5) is not due to the looseness of the bound. It can be shown that as soon as the essential infimum of the optimal arm's distribution $\tilde{\mu} = \sup\{v \in \mathbb{R} : \nu_{i^*}([0, v)) = 0\}$ is smaller than the mean reward of the second best arm, the pseudo-regret admits a polynomial tail only: there exists $C' > 0$ (depending on the distributions $\nu_1, \ldots, \nu_K$) such that for any $C > 0$, there exists $n_0 > 0$ such that for any $n \geq n_0$, $\mathbb{P}\big(\overline{R}_n > C \log n\big) \geq \big(\frac{1}{C' C \log n}\big)^{C'}$. In particular, there is no positive quantities C, C' for which for any n, we have[3] $\mathbb{P}\big(\overline{R}_n > C \log n\big) \leq \frac{C'}{n}$.

The regret concentration, although it improves as ζ grows, is thus pretty slow. The slow concentration happens when the first draws Ω of the optimal arm are unlucky (yielding small rewards) in which case the optimal arm will not be selected any more during roughly the first e^Ω steps. As a result, the distribution of

[2] For instance, when the optimal arm concentrates its rewards on 0 and 1 (Bernoulli distribution with parameter $1/2$), and when the other arms always provide a reward equal to $1/2 - 1/n^{1/6}$, the expected regret is lower bounded by $Cn^{1/7}$.

[3] An entirely analogous result holds for UCB1: using the variance estimates or not does not change the form of the tail distribution of the regret.

the regret can be seen as a mixture of a peaky mode corresponding to situations in which the optimal arm has a "normal" behaviour (with small variations due to the suboptimal arms) and a very thick-tailed mode corresponding to the unlucky start described above. Our theoretical study shows that the mass of this mode decays only at a polynomial rate controlled by ζ. Recall that the larger ζ is, the more all arms are explored, the larger the bound on the expected regret is (see (4.2.4)). In our experiments, this mode does appear (see Figure 4.1).

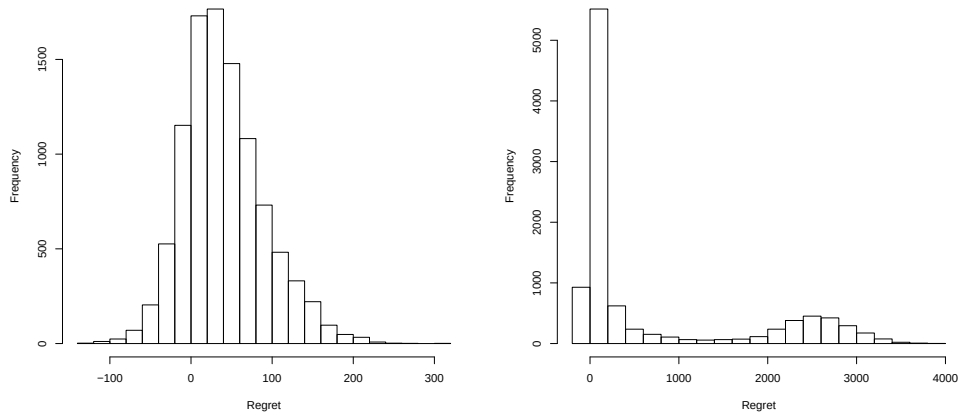


Figure 4.1: Distribution of the pseudo-regret for UCB-V ($\zeta = 1$) for horizon $n = 16,384$ (l.h.s. figure) and $n = 524,288$ (r.h.s. figure). The bandit problem is defined by $K = 2$, a Bernoulli distribution with parameter 0.5 and a Dirac distribution at 0.495.

When the time horizon n is known, one may consider the UCB policy with

$$B_{i,s,t} = \overline{X}_{i,s} + \sqrt{\frac{6\bar{V}_{i,s}\log n}{s}} + \frac{9\log n}{s}, \tag{4.2.6}$$

which is an upper bound on μ_i which holds with probability at least $1 - n^{-3}$. The associated UCB policy, called hereafter UCB-Horizon, concentrates its exploration phase at the beginning of the plays, and then switches to the exploitation mode. On the contrary, the UCB-V induced by (4.2.2), which looks deceptively similar to UCB-Horizon (with $\zeta = 3$), explores and exploits at any time during the interval $[1, n]$. Both policies have similar guarantee on their expected regret. However, on the one hand, UCB-Horizon always satisfies

$$\mathbb{P}\big(\overline{R}_n > C\log n\big) \leq \frac{C'}{n}, \tag{4.2.7}$$

where C and C' are quantities depending only on $K, \zeta, \sigma_1, \ldots, \sigma_K, \Delta_1, \ldots, \Delta_K$, which contrasts with the significantly worse tail distribution of UCB-V. On the

other hand, unlike UCB-Horizon, UCB-V has the anytime property: the policy satisfies the logarithmic expected regret bound for any time horizon n (since its pulling strategy does not depend on the time horizon). The open question here is thus: could we have both properties? In other words, is there an algorithm that does not need to know the time horizon and which regret has a tail distribution satisfying (4.2.7)? We conjecture that the answer is no.

4.2.6. DISTRIBUTION-FREE OPTIMAL UCB POLICY. The inequalities (4.2.3) and (4.2.4) may have surprised the reader since the right-hand sides diverge for Δ_i going to 0. For $\Delta_i = \mathrm{o}(n^{-1/2})$, this is an artefact of the bounds, which is easily rectifiable. For instance, for UCB1, the more general bound (but less readable one) is

$$\mathbb{E}\overline{R}_n \leq \max_{t_i \geq 0, \sum_i t_i = n} \sum_{i:\Delta_i > 0} \min\left(\frac{10}{\Delta_i}\log n\,,\; t_i \Delta_i\right).$$

In the worst case $\big($i.e., $\Delta_1 = 0$ and $\Delta_2 = \cdots = \Delta_K = \sqrt{10K(\log n)/n}\big)$, the right-hand side of the bound is equal to $\sqrt{10n(K-1)\log n}$. This has to be compared with the following lower bound of Auer, Cesa-Bianchi, Freund and Schapire [19]:

$$\inf \sup \mathbb{E}\overline{R}_n \geq \frac{1}{20}\sqrt{nK},$$

where the infimum is taken over all policies and the supremum is taken over all K-tuple of probability distributions on $[0,1]$. We thus observe a logarithmic gap. In [11, 12], Sbastien Bubeck and I close this logarithmic gap, by using a different UCB policy based on

$$B_{i,s,t} = \overline{X}_{i,s} + \sqrt{\frac{\log \max(\frac{n}{Ks}, 1)}{s}},$$

which, for $s < n/K$, is an upper bound on μ_i which holds with probability at least $1 - (Ks/n)^{-2}$ according to Hoeffding's inequality. In this policy, an arm that has been drawn more than n/K times has an index equal to the empirical mean of the rewards obtained from the arm, and when it has been drawn close to n/K times, the logarithmic term is much smaller than the one of UCB1, implying less exploration of this already intensively drawn arm. For this policy, we prove

THEOREM 20 *For* $\Delta = \min_{i \in \{1,\ldots,K\}:\Delta_i > 0} \Delta_i$*, the above policy satisfies*

$$\overline{R}_n \leq \frac{23K}{\Delta}\log\left(\max\left(\frac{110n\Delta^2}{K}, 10^4\right)\right), \tag{4.2.8}$$

and

$$\mathbb{E}\overline{R}_n \leq 24\sqrt{nK}. \tag{4.2.9}$$

This means that this UCB policy has the minimax rate $\sqrt{nK}$, while still having a distribution-dependent bound increasing logarithmically in n.

4.2.7. UCB POLICY WITH AN INFINITE NUMBER OF ARMS. When the number of arms is infinite (or larger than the available number of experiments), the exploration of all the arms is impossible: if no additional assumption is made, it may be arbitrarily hard to find a near-optimal arm. In [129], Yizao Wang, Rmi Munos and I consider a stochastic assumption on the *mean-reward* of any new selected arm. When a new arm i is pulled, its mean-reward μ_i is assumed to be an independent sample from a fixed distribution. Our assumptions essentially characterize the probability of pulling near-optimal arms. That is, given $\mu^* \in [0,1]$ as the best possible mean-reward and $\beta \geq 0$ a parameter of the mean-reward distribution, the probability that a new arm is δ-optimal is of order δ^β for small δ, i.e. $\mathbb{P}(\mu_k \geq \mu^* - \delta) = \Theta(\delta^\beta)$ for $\delta \to 0$[4]. In contrast with the previous many-armed bandits [30, 121], our setting allows general reward distributions for the arms, under a simple assumption on the mean-reward.

When there is more arms than the available number of experiments, the exploration takes two forms: discovery (pulling a new arm that has never been tried before) and sampling (pulling an arm already discovered in order to gain information about its actual mean-reward).

Numerous applications can be found e.g. in [30]. It includes labor markets (a worker has many opportunities for jobs), mining for valuable resources (such as gold or oil) when there are many areas available for exploration (the miner can move to another location or continue in the same location, depending on results), and path planning under uncertainty in which the path planner has to decide among a route that has proved to be efficient in the past (exploitation), or a known route that has not been explored many times (sampling), or a brand new route that has never been tried before (discovery).

In [129], we propose an arm-increasing rule policy. It has the anytime property and consists in adding a new arm from time to time into the set of sampled arms. It is done such that at time t, the number of sampled arms is of order $n^{\beta/2}$ if $\mu^* < 1$ and $\beta < 1$, and of order $n^{\beta/(1+\beta)}$ otherwise. It uses a modified version of the UCB-V policy on this set of arms: specifically, the policy associated with

$$B_{i,s,t} = \overline{X}_{i,s} + \sqrt{\frac{4\bar{V}_{i,s}\log(10\log t)}{s}} + \frac{6\log(10\log t)}{s}.$$

The pseudo-regret of this policy is still defined as the difference between the rewards we would have obtained by drawing an optimal arm (an arm having a

[4] We write $f(\delta) = \Theta(g(\delta))$ for $\delta \to 0$ when $\exists c_1, c_2, \varepsilon_0 > 0$ such that $\forall \delta \leq \delta_0$, $c_1 g(\delta) \leq f(\delta) \leq c_2 g(\delta)$.

mean-reward equal to μ^*) and the rewards we did obtain during the time steps $1, \dots, n$, hence, from the tower rule, $\mathbb{E}\overline{R}_n = n\mu^* - \sum_{t=1}^{n} \mu_{I_t}$. Its behaviour depends on whether $\mu^* = 1$ or $\mu^* < 1$. Let us write $v_n = \tilde{O}(u_n)$ when for some $n_0, C > 0$, $v_n \leq C u_n (\log(u_n))^2$, for all $n \geq n_0$. For $\mu^* = 1$, our algorithms are such that $\mathbb{E}\overline{R}_n = \tilde{O}(n^{\beta/(1+\beta)})$. For $\mu^* < 1$, we have $\mathbb{E}\overline{R}_n = \tilde{O}(n^{\beta/(1+\beta)})$ if $\beta > 1$, and (only) $\mathbb{E}\overline{R}_n = \tilde{O}(n^{1/2})$ if $\beta \leq 1$. Moreover we derive the lower bound: for any $\beta > 0$, $\mu^* \leq 1$, any algorithm satisfies $\mathbb{E}\overline{R}_n \geq C n^{\beta/(1+\beta)}$ for some $C > 0$.

In continuum-armed bandits (see e.g. [1, 78, 20]), an infinity of arms is also considered. The arms lie in some Euclidean (or metric) space and their mean-reward is a deterministic and smooth (e.g. Lipschitz) function of the arms. This setting is different from ours since our assumption is stochastic and does not consider regularities of the mean-reward w.r.t. the arms. However, if we choose an arm-pulling strategy which consists in selecting randomly the arms, then our setting encompasses continuum-armed bandits. For example, consider the domain $[0,1]^d$ and a mean-reward function μ assumed to be locally equivalent to a Hölder function (of order $\alpha \in [0, +\infty)$) around any maximum x^* (the number of maxima is assumed to be finite), i.e.

$$\mu(x^*) - \mu(x) = \Theta(\|x^* - x\|^\alpha) \text{ when } x \to x^*. \tag{4.2.10}$$

Pulling randomly an arm X according to the Lebesgue measure on $[0,1]^d$, we have: $\mathbb{P}(\mu(X) > \mu^* - \varepsilon) = \Theta(\mathbb{P}(\|X - x^*\|^\alpha < \varepsilon)) = \Theta(\varepsilon^{d/\alpha})$, for $\varepsilon \to 0$. Thus our assumption holds with $\beta = d/\alpha$, and our results say that if $\mu^* = 1$, we have $\mathbb{E}\overline{R}_n = \tilde{O}(n^{\beta/(1+\beta)}) = \tilde{O}(n^{d/(\alpha+d)})$.

For $d = 1$, under the assumption that μ is α-Hölder (i.e. $|\mu(x) - \mu(y)| \leq c\,\|x - y\|^\alpha$ for $0 < \alpha \leq 1$), [78] provides upper and lower bounds on the pseudo-regret $\overline{R}_n = \Theta(n^{(\alpha+1)/(2\alpha+1)})$. Our results gives $\mathbb{E}\overline{R}_n = \tilde{O}(n^{1/(\alpha+1)})$ which is better for all values of α. The reason for this apparent contradiction is that the lower bound in [78] is obtained by the construction of a very irregular function, which actually does not satisfy our local assumption (4.2.10).

Now, under assumptions (4.2.10) for any $\alpha > 0$ (around a finite set of maxima), [20] provides the rate $\mathbb{E}\overline{R}_n = \tilde{O}(\sqrt{n})$. Our result gives the same rate when $\mu^* < 1$ but in the case $\mu^* = 1$ we obtain the improved rate $\mathbb{E}\overline{R}_n = \tilde{O}(n^{1/(\alpha+1)})$ which is better whenever $\alpha > 1$ (because we are able to exploit the low variance of the good arms). Note that like our algorithm, the algorithms in [20] as well as in [78], do not make an explicit use (in the procedure) of the smoothness of the function. They just use a "uniform" discretization of the domain.

On the other hand, the zooming algorithm of [75] adapts to the smoothness of μ (more arms are sampled at areas where μ is high). For any dimension d, they obtain $\mathbb{E}\overline{R}_n = \tilde{O}(n^{(d'+1)/(d'+2)})$, where $d' \leq d$ is their "zooming dimension". Under assumptions (4.2.10) we deduce $d' = \frac{\alpha-1}{\alpha} d$ using the Euclidean distance as

metric, thus their pseudo-regret is $\mathbb{E}\overline{R}_n = \tilde{O}(n^{(d(\alpha-1)+\alpha)/(d(\alpha-1)+2\alpha)})$. For locally quadratic functions (i.e. $\alpha = 2$), their rate is $\tilde{O}(n^{(d+2)/(d+4)})$, whereas ours is $\tilde{O}(n^{d/(2+d)})$. Again, we have a smaller pseudo-regret although we do not use the smoothness of μ in our algorithm. Here the reason is that the zooming algorithm does not make full use of the fact that the function is locally quadratic (it considers a Lipschitz property only). However, in the case $\alpha < 1$, our rates are worse than algorithms specifically designed for continuum armed bandits.

Hence, the comparison between the many-armed and continuum-armed bandits settings is not easy because of the difference in nature of the basis assumptions. Our setting is an alternative to the continuum-armed bandit setting which does not require the existence of an underlying metric space in which the mean-reward function would be smooth. Our assumption naturally deals with possibly very complicated functions where maxima may be located in any part of the space. For the continuum-armed bandit problems when there are relatively many near-optimal arms, our algorithm will be also competitive compared to the specifically designed continuum-armed bandit algorithms. This result matches the intuition that in such cases, a random selection strategy will perform well.

Another contribution of our work is to show that, for infinitely many-armed bandits, we need much less exploration of each arm than for finite-armed bandits: as shown in the next section, the index $B_{i,s,t}$ is an upper bound on μ_i which holds with probability at least $1 - [\log(10t)]^{-2}$. The use of this low confidence upper bound (compared to the ones of UCB1 and UCB-V for instance) can be explained by the fact that many sampled arms have a mean really close to the optimal one, and consequently exploiting not the best one but just one of the best arms is enough to achieve the minimax pseudo-regret.

4.2.8. THE EMPIRICAL BERNSTEIN INEQUALITY. A key lemma to analyze the policies using variance estimates as UCB-V and the one used in the previous section is the following maximal inequality, which in particular implies that the arm index (4.2.2) of UCB-V is an upper bound on μ_i which holds with probability at least $1 - 3t^{-\zeta}$. The interest of the lemma goes beyond the particular setting of the multi-armed bandit problems as it provides a *non asymptotic* confidence interval on the expectation of a distribution for which we observe a sample (and for which we know a bounded interval containing its support).

LEMMA 21 *Let $U, U_1, \ldots, U_n$ be independent and identically distributed random variables taking their values in $[0,1]$. Let*

$$\bar{U}_t = \frac{1}{t}\sum_{i=1}^{t} U_i \qquad \textit{and} \qquad \bar{V}_t = \frac{1}{t}\sum_{i=1}^{t}(U_i - \bar{U}_t)^2.$$

1. *For any $\varepsilon > 0$, with probability at least $1-\varepsilon$, for any $t \in \{1,\dots,n\}$ and $\ell_t = \frac{n\log(2\varepsilon^{-1})}{t^2}$, we have*

$$\bar{U}_t - \mathbb{E}U < \min\left(\sqrt{2\ell_t(\bar{V}_t+\ell_t)} + \ell_t\Big(\frac{1}{3}+\sqrt{1-3\bar{V}_t}\Big)\ ,\ \sqrt{\frac{\ell_t}{2}}\right). \quad (4.2.11)$$

2. *For any $\varepsilon > 0$, with probability at least $1-\varepsilon$, for any $t \in \{1,\dots,n\}$ and $\tilde{\ell}_t = \frac{n\log(3\varepsilon^{-1})}{t^2}$, we have*

$$\big|\bar{U}_t - \mathbb{E}U\big| < \min\left(\sqrt{2\tilde{\ell}_t(\bar{V}_t+\tilde{\ell}_t)} + \tilde{\ell}_t\Big(\frac{1}{3}+\sqrt{1-3\bar{V}_t}\Big)\ ,\ \sqrt{\frac{\tilde{\ell}_t}{2}}\right). \quad (4.2.12)$$

In particular, for any $\varepsilon > 0$, with probability at least $1-\varepsilon$, for any $t \in \{1,\dots,n\}$, we have

$$\big|\bar{U}_t - \mathbb{E}U\big| < \sqrt{\frac{2n\bar{V}_t\log(3\varepsilon^{-1})}{t^2}} + \frac{3n\log(3\varepsilon^{-1})}{t^2}. \quad (4.2.13)$$

Inequality (4.2.13) is the one used in [15, 109], but its tighter version (4.2.12) should be preferred. The proof of this lemma is given in Appendix E. Fot $t = n$, the lemma is an empirical version of Bernstein's inequality, which differs from the latter to the following extent: the true variance has been replaced by its empirical estimate (at the price of having $\log(3\varepsilon^{-1})$ terms instead of $\log(\varepsilon^{-1})$, and a factor 3 in the last term in the right-hand side instead of $1/3$. Inequality (4.2.13) relies on the following empirical upper bound of the variance V of U, which simultaneously holds with probability at least $1-\varepsilon$: for any $t \in \{1,\dots,n\}$, we have

$$V \le \left(\sqrt{\bar{V}_t + \frac{n\log(3\varepsilon^{-1})}{t^2}} + \sqrt{\frac{n\log(3\varepsilon^{-1})}{2t^2}(1-3\bar{V}_t)}\right)^2.$$

This bound can be seen as an improvement of Inequality (5.27) of Blanchard [33]. For $t = n \ge 2$, i.e. without the stopping time argument due to Freedman [57] allowing to have the inequality uniformly over time, Maurer and Pontil [101] improves on the constants of the above inequality when the empirical variance is close to 0. Considering the unbiased variance estimator $\bar{V}_t' = \frac{1}{t-1}\sum_{s=1}^t (U_s - \bar{U}_t)^2 = \frac{t}{t-1}\bar{V}_t$, they obtain that with probability at least $1-\varepsilon$,

$$V \le \left(\sqrt{\bar{V}_t' + \frac{\log(\varepsilon^{-1})}{2(t-1)}} + \sqrt{\frac{\log(\varepsilon^{-1})}{2(t-1)}}\right)^2.$$

Combined with Bernstein's bound, this gives that with probability at least $1-\varepsilon$,

$$\left|\bar{U}_t - \mathbb{E}U\right| \leq \sqrt{\frac{2\log(3\varepsilon^{-1})}{t}\left(\bar{V}_t' + \frac{\log(3\varepsilon^{-1})}{2(t-1)}\right)} + \frac{4\log(3\varepsilon^{-1})}{3(t-1)},$$

where the gain is on the factor of the logarithmic term when the empirical variance is much smaller than $\log(3\varepsilon^{-1})/t$.

Volodymyr Mnih, Csaba Szepesvári and I [109] have used Lemma 21 to address the problem of stopping the sampling of an unknown distribution ν as soon as we can output an estimate $\hat{\mu}$ of the mean μ of ν with relative error δ with probability at least $1-\varepsilon$, that is

$$\mathbb{P}\big(|\hat{\mu}-\mu| \leq \delta|\mu|\big) \geq 1-\varepsilon, \tag{4.2.14}$$

For a distribution ν supported by $[a, a+1]$ for some $a \in \mathbb{R}$, we have proposed the empirical Bernstein stopping algorithm described in Figure 4.2. It uses a geometric grid and parameters ensuring that the event $\mathcal{E} = \{\left|\bar{U}_t - \mu\right| \leq c_t,\, t \geq t_1\}$ occurs with probability at least $1-\varepsilon$. It operates by maintaining a lower bound, LB, and an upper bound, UB, on the absolute value of the mean of the random variable being sampled, terminates when $(1+\delta)\text{LB} < (1-\delta)\text{UB}$, and returns the mean estimate $\hat{\mu} = \operatorname{sign}(\bar{U}_t)\frac{(1+\delta)\text{LB}+(1-\delta)\text{UB}}{2}$. We prove that this output indeed satisfies (4.2.14) and that the stopping time T of the algorithm is upper bounded by

$$T \leq C \cdot \max\left(\frac{\sigma^2}{\delta^2\mu^2}, \frac{1}{\delta|\mu|}\right)\left(\log\left(\frac{2}{\varepsilon}\right) + \log\left(\log\frac{3}{\delta|\mu|}\right)\right).$$

Up to the $\log\log$ term, this is optimal according to the work of Dagum, Karp, Luby and Ross [49].

Besides, our experimental simulations show that it significantly outperforms previously known stopping rules, in particular AA [49] and the Nonmonotonic Adaptive Sampling (NAS) algorithm due to Domingo, Gavalda and Watanabe [130, 54]. Figure 4.3 shows the results of running different stopping rules for the distribution ν of the average of 10 uniform random variables on $[\mu-1/2, \mu+1/2]$ with varying μ and also on Bernoulli distributions. The experience is repeated a hundred times so that the differences observed in Figure 4.3 are statistically significant.

We also use the empirical Bernstein bound in the context of racing algorithms. Racing algorithms aim to reduce the computational burden of performing tasks such as model selection using a hold-out set by discarding poor models quickly [98, 112]. The context of racing algorithms is the one of multi-armed bandit problems. Let $\varepsilon > 0$ be the confidence level parameter. A racing algorithm either terminates when it runs out of time (i.e. at the end of the n-th round) or when it

Parameters of the problem: δ, ε and the unknown distribution ν.
Parameters of the algorithm: $q > 0$, $t_1 \geq 1$ and $\alpha > 1$ defining the geometric grid $t_k = \lceil \alpha t_{k-1} \rceil$. (In our simulations, we take $q = 0.1$, $t_1 = 20$ and $\alpha = 1.1$.)

Initialization:
$c = \frac{3}{\varepsilon t_1^q (1-\alpha^{-q})}$
LB $\leftarrow 0$
UB $\leftarrow \infty$

For $t = 1, \ldots, t_1 - 1$,
  sample U_t from ν
End For

For $k = 1, 2, \ldots,$
  For $t = t_k, \ldots, t_{k+1} - 1$,
    sample U_t from ν and compute the empirical mean $\bar{U}_t = \frac{1}{t} \sum_{s=1}^{t} U_s$
    $\ell_t = \frac{t_{k+1}}{t^2} \log(c t_k^q)$.
    $c_t = \min \left(\sqrt{2\ell_t(\bar{V}_t + \ell_t)} + \ell_t \left(\frac{1}{3} + \sqrt{1 - 3\bar{V}_t} \right), \sqrt{\frac{\ell_t}{2}} \right)$
    LB $\leftarrow$ max(LB, $|\bar{U}_t| - c_t$)
    UB $\leftarrow$ min(UB, $|\bar{U}_t| + c_t$)
    If $(1+\delta)$LB $< (1-\delta)$UB, Then
      stop simulating U and return the mean estimate $\mathrm{sign}(\bar{U}_t) \frac{(1+\delta)\text{LB} + (1-\delta)\text{UB}}{2}$
    End If
  End For
End For

Figure 4.2: Empirical Bernstein stopping (EBGStop* in our experiments).

can say that with probability at least $1 - \varepsilon$, it has found the best option, i.e. an option $i^* \in \text{argmax}_{i \in \{1,\ldots,K\}} \mu_i$.

The Hoeffding race introduced by [98] is an algorithm based on discarding options which are likely to have smaller mean than the optimal one until only one option remains. Precisely, for each time step and each distribution, $\frac{\delta}{nK}$-confidence intervals are constructed for the mean. Options with upper confidence smaller than the lower confidence bound of another option are discarded. The algorithm samples one by one all the options that have not been discarded yet. Our empirical and theoretical study show that replacing the Hoeffding's inequality by the empirical Bernstein bound leads to significant improvement. In particular, Table 4.1 shows the percentage of work saved by each method ($1-$ number of samples taken by method divided by Kn), as well as the number of options remaining after termination (see [109] for a more detailed description of the experiments).

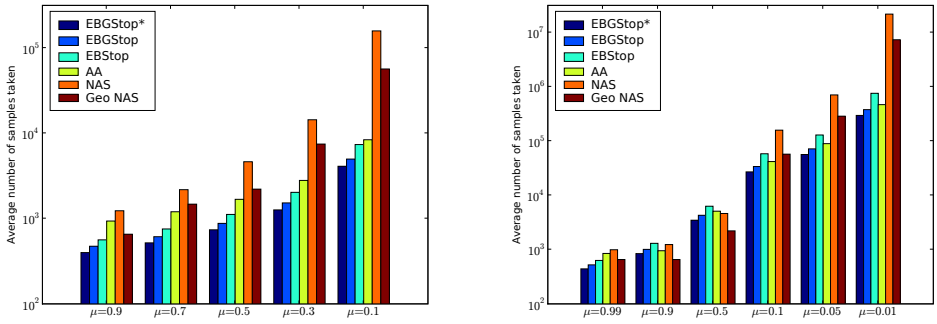


Figure 4.3: Comparison of stopping rules on (l.h.s. figure) averages of uniform random variables with varying means and (r.h.s. figure) Bernoulli random variables with means 0.99, 0.9, 0.5, 0.1, 0.05, and 0.01, averaged over 100 runs. The average number of samples is shown in log scale.

Table 4.1: Percentage of work saved / number of options left after termination.

Data set	Hoeffding	Empirical Bernstein
SARCOS	0.0% / 11	44.9% / 4
Covertype2	14.9% / 8	29.3% / 5
Local	6.0% / 9	33.1% / 6

4.2.9. BEST ARM IDENTIFICATION. Racing algorithms [98] try to identify the best action at a given confidence level while consuming the minimal number of pulls. They essentially try to optimize the exploration "budget" for a given confidence level. In some applications, the budget size is fixed (say n rounds), and one may want to predict the best arm at the end of the n-th round. A motivating example concerns channel allocation for mobile phone communications. During a very short time before the communication starts, a cellphone can explore the set of channels to find the best one to operate. Each evaluation of a channel is noisy and there is a limited number of evaluations before the communication starts. The connection is then launched on the channel which is believed to be the best.

More formally, the setting of identifying the best arm is summarized in Figure 4.4. It differs from the traditional multi-armed bandit problem by its target: the cumulative regret is no longer appropriate to measure the performance of a policy. The aim is rather to minimize the simple regret:

$$r_n = \Delta_{J_n},$$

where J_n is the final recommendation of the algorithm and Δ_i still denotes the gap between the mean reward of the best arm or the mean reward of the selected arm. Let i^* still denote the optimal arm. The simple regret is linked to the probability

Parameters available to the forecaster: the number of rounds n and the number of arms K.

Parameters unknown to the forecaster: the reward distributions $\nu_1, \ldots, \nu_K$ of the arms.

For each round $t = 1, 2, \ldots, n$;

(1) the forecaster chooses $I_t \in \{1, \ldots, K\}$,

(2) the environment draws the reward $X_{I_t, T_{I_t}(t)}$ from ν_{I_t} and independently of the past given I_t.

At the end of the n rounds, the forecaster outputs a recommendation $J_n \in \{1, \ldots, K\}$.

Figure 4.4: Best arm identification in multi-armed bandits.

of error

$$e_n = \mathbb{P}(J_n \neq i^*),$$

since, from $\mathbb{E}r_n = \sum_{i \neq i^*} \mathbb{P}(J_n = i)\Delta_i$, we have $\min_{i:\Delta_i>0} \Delta_i e_n \leq \mathbb{E}r_n \leq e_n$.

In [13], Sbastien Bubeck, Rmi Munos and I prove that UCB policies can still be used provided that the exploration term is taken much larger: precisely, for $H = \sum_{i:\Delta_i>0} \Delta_i^{-2}$ and a numerical constant $c > 0$, we introduce the UCB-E (E for exploration) policy characterized by

$$B_{i,s,t} = \overline{X}_{i,s} + \sqrt{\frac{cn}{2sH}},$$

which is an extremely high confidence upper bound on μ_i (probability at least $1 - \exp(-\frac{cn}{H})$, hence much higher than the confidence level of UCB1 and UCB-V), and by taking J_n as the arm with the largest empirical mean. We also propose a new algorithm, called SR, based on successive rejects. We show that these algorithms are essentially optimal since their simple regret decreases exponentially at a rate which is, up to a logarithmic factor, the best possible. However, while the UCB policy needs the tuning of a parameter depending on the unobservable hardness of the task, the successive rejects policy benefits from being parameter-free, and also independent of the scaling of the rewards. As a by-product of our analysis, we show that identifying the best arm (when it is unique) requires a number of samples of order H (up to a $\log(K)$ factor). This generalizes the well-known

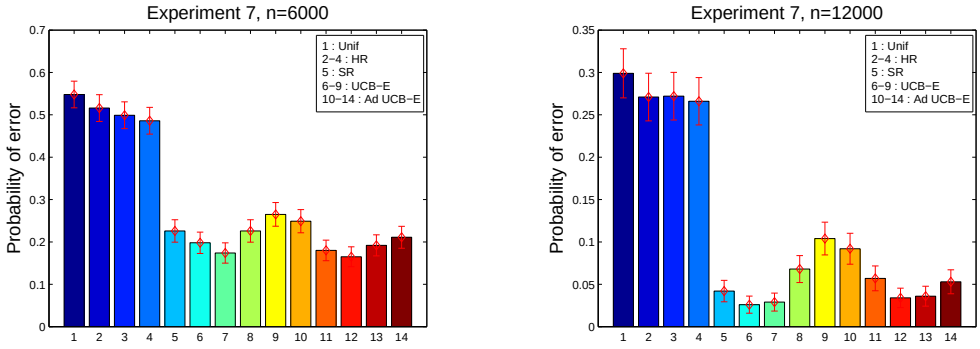


Figure 4.5: Probability of error of different algorithms for $n = 6000$ (l.h.s.) and $n = 12000$ (r.h.s.), and $K = 30$ arms having Bernoulli distributions with parameters 0.5 (one arm), 0.45 (five arms), 0.43 (fourteen arms), 0.38 (ten arms). Each bar represents a different algorithm and the bar's height represents the probability of error of this algorithm. "Unif" is the uniform sampling strategy, "HR" is the Hoeffding Race algorithm (run for three different values of the confidence level parameter), UCB-E is tested for four different values of c: $2, 4, 8, 16$, Adaptive UCB-E is tested for five different values of its parameter. More extensive experiments are presented in [13] and confirm the ranking of algorithms observed on these simulations: Ad UCB-E ¿ SR ¿ HR ¿ Unif, where '¿' means 'has better performance than'. (UCB-E is not ranked as it requires the knowledge of H.)

fact that one needs of order of $1/\Delta^2$ samples to differentiate the means of two distributions with gap Δ. A precise understanding of both SR and the UCB-E policy leads us to define a new algorithm, Adaptive UCB-E. It comes without guarantee of optimal rates, but performs slightly better than SR in practice as shown in Figure 4.5.

Another variant of the best arm identification task is the problem of minimal sampling times required to identify an ϵ-optimal arm with a given confidence level, see in particular [54] and [56]. In [62], Steffen Grünewälder, Manfred Opper, John Shawe-Taylor and I also study a non-cumulative regret notion, but in the context of a continuum of arms. Precisely, we consider the scenario in which the reward distribution for arms is modelled by a Gaussian process and there is no noise in the observed reward, and provide upper and lower bounds under reasonable assumptions about the covariance function defining the Gaussian process.

4.3. SEQUENTIAL PREDICTION

This section summarizes my work with Sbastien Bubeck [12]. It starts with the adversarial bandit problem, and goes on with the extension to other sequential

prediction games.

4.3.1. ADVERSARIAL BANDIT. In the general bandit problem, the environment is not constrained to generate the reward vectors independently as in the stochastic bandit problem. However, the target is still to minimize the regret

$$R_n = \max_{i=1,\dots,K} \sum_{t=1}^{n} \big(g_{i,t} - g_{I_t,t}\big).$$

In the most general form of the game, called the non-oblivious/adaptive adversarial game, the adversary may choose the reward vector g_t as a function of the past decisions $I_1, \dots, I_{t-1}$. Upper bounds on the regret R_n for this type of adversary have a less straightforward interpretation since the target cumulative reward is now depending on the agent's policy! I will not provide here results for this type of adversary but the extension of the results presented hereafter can be found in [12].

Thus we will focus on the oblivious adversarial bandit game, in which the reward vector g_t is *not* a function of the past decisions $I_1, \dots, I_{t-1}$. The environment is then simply defined by a distribution on $[0,1]^{nK}$, while the agent's policy is still defined by a mapping, denoted φ from $\cup_{t\in\{1,\dots,n-1\}} \big(\{1,\dots,K\} \times [0,1]\big)^t$ to the set of distributions of $\{1,\dots,K\}$. Now we can see the game a bit differently. The "master" of the game draws a matrix $(g_{i,t})_{1\le i\le K, 1\le t\le n}$ from the distribution defining the environment, and at each time step t, draws the arm I_t according to the distribution $p_t = \varphi(\mathcal{H}_t)$ chosen by the agent, where $\mathcal{H}_t = \{(I_1, g_{I_1,1}), \dots, (I_{t-1}, g_{I_{t-1},t-1})\}$ is the past information. The regret R_n is a random variable since it depends on the draw of the reward matrix and the draws from the distributions p_t's.

In [19], Auer, Cesa-Bianchi, Freund and Schapire have shown that a forecaster based on exponentially weighted averages has a regret upper bounded by $2.7\sqrt{nK\log K}$. As stated before, they also show that this is optimal up to the logarithmic factor: precisely, there is no forecaster satisfying $\mathbb{E}\overline{R}_n \le \frac{1}{20}\sqrt{nK}$, for any environment. In [11, 12], we close the logarithmic gap between the above upper and lower bounds by introducing a new class of randomized policies. To define it, consider a function $\psi : \mathbb{R}_-^* \to \mathbb{R}_+^*$ such that

$$\begin{array}{l} \psi \text{ increasing and continuously differentiable,} \\ \psi'/\psi \text{ nondecreasing,} \\ \lim_{u\to-\infty} \psi(u) < 1/K, \text{ and } \lim_{u\to 0} \psi(u) \ge 1. \end{array} \tag{4.3.1}$$

It can be easily shown that there exists a continuously differentiable function $C : \mathbb{R}_+^K \to \mathbb{R}$ satisfying for any $x = (x_1, \dots, x_K) \in \mathbb{R}_+^K$,

$$\max_{i=1,\dots,K} x_i < C(x) \le \max_{i=1,\dots,K} x_i - \psi^{-1}(1/K), \tag{4.3.2}$$

Parameter: function $\psi : \mathbb{R}_-^* \to \mathbb{R}_+^*$ satisfying (4.3.1)
Let p_1 be the uniform distribution over $\{1, \ldots, K\}$.

For each round $t = 1, 2, \ldots,$

(1) Draw an arm I_t from the probability distribution p_t.

(2) Compute the estimated gain for each arm: $\tilde{g}_{i,t} = \frac{g_{i,t}}{p_{i,t}} \mathbb{1}_{I_t=i}$ and[a] update the estimated cumulative gain: $\tilde{G}_{i,t} = \sum_{s=1}^{t} \tilde{g}_{i,s}$.

(3) Compute the normalization constant $C_t = C(\tilde{G}_t)$ where $\tilde{G}_t = (\tilde{G}_{1,t}, \ldots, \tilde{G}_{K,t})$.

(4) Compute the new probability distribution $p_{t+1} = (p_{1,t+1}, \ldots, p_{K,t+1})$ where

$$p_{i,t+1} = \psi(\tilde{G}_{i,t} - C_t).$$

[a] It estimates $g_{i,t}$ even when $g_{i,t}$ is not observed since $\mathbb{E}\tilde{g}_{i,t} = g_{i,t}$.

Figure 4.6: INF (Implicitly Normalized Forecaster) for the adversarial bandit.

and

$$\sum_{i=1}^{K} \psi(x_i - C(x)) = 1. \tag{4.3.3}$$

So we can define the implicitly normalized forecaster (INF) as detailed in Figure 4.6. Indeed, Equality (4.3.3) makes the fourth step in Figure 4.6 legitimate. From (4.3.2), $C(\tilde{G}_t)$ is roughly equal to $\max_{i=1,\ldots,K} \tilde{G}_{i,t}$. This means that INF chooses the probability assigned to arm i as a function of the (estimated) regret. In spirit, it is similar to the traditional weighted average forecaster, see e.g. Section 2.1 of [46], where the probabilities are proportional to a function of the difference between the (estimated) cumulative reward of arm i and the cumulative reward of the policy, which should be, for a well-performing policy, of order $C(\tilde{G}_t)$. Weigthed average forecasters and implicitly normalized forecasters are in fact two different classes of forecasters which intersection contains exponentially weighted average forecasters such as the one considered in [19]. The interesting feature of the implicit normalization is the following argument, which allows to recover the result of [19] and more interestingly to propose a policy having a regret of order $\sqrt{nK}$. It starts with an Abel transformation and consequently is "orthogonal" to the usual argument which, for sake of comparison, has been reproduced in

Appendix F.2. Letting $\tilde{G}_0 = 0 \in \mathbb{R}^K$. We have

$$\begin{aligned}
\sum_{t=1}^{n} g_{I_t,t} &= \sum_{t=1}^{n}\sum_{i=1}^{K} p_{i,t}\tilde{g}_{i,t} \\
&= \sum_{t=1}^{n}\sum_{i=1}^{K} p_{i,t}(\tilde{G}_{i,t} - \tilde{G}_{i,t-1}) \\
&= \sum_{i=1}^{K} p_{i,n+1}\tilde{G}_{i,n} + \sum_{i=1}^{K}\sum_{t=1}^{n} \tilde{G}_{i,t}(p_{i,t} - p_{i,t+1}) \\
&= \sum_{i=1}^{K} p_{i,n+1}\left(\psi^{-1}(p_{i,n+1}) + C_n\right) + \sum_{i=1}^{K}\sum_{t=1}^{n}(\psi^{-1}(p_{i,t+1}) + C_t)(p_{i,t} - p_{i,t+1}) \\
&= C_n + \sum_{i=1}^{K} p_{i,n+1}\psi^{-1}(p_{i,n+1}) + \sum_{i=1}^{K}\sum_{t=1}^{n} \psi^{-1}(p_{i,t+1})(p_{i,t} - p_{i,t+1}),
\end{aligned} \tag{4.3.4}$$

where the remarkable simplification in the last step is closely linked to our specific class of randomized algorithms. The equality is interesting since, from (4.3.2), C_n approximates the maximum estimated cumulative reward $\max_{i=1,\dots,K} \tilde{G}_{i,n}$, which should be close to the cumulative reward of the optimal arm $\max_{i=1,\dots,K} G_{i,n}$, where $G_{i,n} = \sum_{t=1}^{n} g_{i,t}$. Besides, the last term in the right-hand side is roughly equal to

$$\sum_{i=1}^{K}\sum_{t=1}^{n}\int_{p_{i,t}}^{p_{i,t+1}} \psi^{-1}(u)du = \sum_{i=1}^{K}\int_{1/K}^{p_{i,n+1}} \psi^{-1}(u)du$$

To make this precise, we use a Taylor-Lagrange expansion and technical arguments to control the residual terms. Putting this together, we roughly have

$$\max_{i=1,\dots,K} G_{i,n} - \sum_{t=1}^{n} g_{I_t,t} \lessapprox -\sum_{i=1}^{K} p_{i,n+1}\psi^{-1}(p_{i,n+1}) + \sum_{i=1}^{K}\int_{1/K}^{p_{i,n+1}} \psi^{-1}(u)du.$$

The right-hand side is easy to study: it depends only on the final probability vector and has simple upper bounds for adequate choices of ψ. For instance, for $\psi(x) = \exp(\eta x) + \frac{\gamma}{K}$ with $\eta > 0$ and $\gamma \in [0,1)$, the right-hand side is smaller than $\frac{1-\gamma}{\eta}\log\left(\frac{K}{1-\gamma}\right) + \gamma C_n$. For $\psi(x) = \left(\frac{\eta}{-x}\right)^q + \frac{\gamma}{K}$ with $\eta > 0$, $q > 1$ and $\gamma \in [0,1)$, it is smaller than $\frac{q}{q-1}\eta K^{1/q} + \gamma C_n$. For sake of simplicity, we have been hiding the residual terms but these terms when added together (nK terms!) are not that small, and in fact constrain the choice of the parameters γ and η if one wishes to get the tightest bound. Our main result is the following.

Parameters: the number of arms (or actions) K and the number of rounds n with $n \geq K \geq 2$.

For each round $t = 1, 2, \ldots, n$

(1) The forecaster chooses an arm $I_t \in \{1, \ldots, K\}$, possibly with the help of an external randomization.

(2) Simultaneously the adversary chooses the reward vector

$$g_t = (g_{1,t}, \ldots, g_{K,t}) \in [0,1]^K$$

(3) The forecaster receives the gain $g_{I_t,t}$ (without systematically observing it). He observes

- the reward vector $(g_{1,t}, \ldots, g_{K,t})$ in the **full information** game,
- the reward vector $(g_{1,t}, \ldots, g_{K,t})$ if he asks for it with the global constraint that he is not allowed to ask it more than m times for some fixed integer number $1 \leq m \leq n$. This prediction game is the **label efficient** game,
- only $g_{I_t,t}$ in the **bandit** game,
- only his obtained reward $g_{I_t,t}$ if he asks for it with the global constraint that he is not allowed to ask it more than m times for some fixed integer number $1 \leq m \leq n$. This prediction game is the **bandit label efficient** game.

Goal : The forecaster tries to maximize his cumulative gain $\sum_{t=1}^{n} g_{I_t,t}$.

Figure 4.7: The four prediction games considered in this section.

THEOREM 22 *The INF algorithm with* $\psi(x) = \left(\frac{3\sqrt{n}}{-x}\right)^2 + \frac{1}{\sqrt{nK}}$ *satisfies*

$$\mathbb{E}R_n \leq 11\sqrt{nK}.$$

4.3.2. EXTENSIONS TO OTHER SEQUENTIAL PREDICTION GAMES. Let us now describe a more general setting, in which the feedback received by the forecaster after drawing an arm differs from game to game. The four games are detailed in Figure 4.7. As for the weighted average forecasters, the INF forecaster can be adapted to the different games by simply modifying the estimates $\tilde{g}_{i,t}$ of $g_{i,t}$. The resulting slightly modified INF forecaster is given in Figure 4.8. Interestingly, we can provide a unified analysis of these games for the INF forecaster. It allows to

essentially recover the known minimax bounds, while sometimes improving the best known upper bound by a logarithmic term. It also leads to high probability bounds on the regret holding for any confidence level, which contrasts with previously known results. Let us now detail the main results for the last three games of Figure 4.8 and for the tracking the best expert scenario.

INF (Implicitly Normalized Forecaster):

Parameters:

- the continuously differentiable function $\psi : \mathbb{R}_-^* \to \mathbb{R}_+^*$ satisfying (4.3.1)
- the estimates $\tilde{g}_{i,t}$ of $g_{i,t}$ based on the (drawn arms and) observed rewards at time t (and before time t)

Let p_1 be the uniform distribution over $\{1, \ldots, K\}$.

For each round $t = 1, 2, \ldots,$

(1) Draw an arm I_t from the probability distribution p_t.

(2) Use the (potentially) observed reward(s) to build the estimate $\tilde{g}_t = (\tilde{g}_{1,t}, \ldots, \tilde{g}_{K,t})$ of $(g_{1,t}, \ldots, g_{K,t})$ and let: $\tilde{G}_t = \sum_{s=1}^t \tilde{g}_s = (\tilde{G}_{1,t}, \ldots, \tilde{G}_{K,t})$.

(3) Compute the normalization constant $C_t = C(\tilde{G}_t)$.

(4) Compute the new probability distribution $p_{t+1} = (p_{1,t+1}, \ldots, p_{K,t+1})$ where

$$p_{i,t+1} = \psi(\tilde{G}_{i,t} - C_t).$$

Figure 4.8: The proposed policy for the four prediction games.

The label efficient game. This game was introduced by [66]: as explained in Figure 4.7, the forecaster observes the reward vector only if he asks for it, and he is not allowed to ask it more than m times for some fixed integer number $1 \leq m \leq n$. Following the work of Cesa-Bianchi, Lugosi and Stoltz [47], we consider the following policy for requesting the reward vector. At each round, we draw a Bernoulli random variable Z_t, with parameter $\delta = \frac{3m}{4n}$, to decide whether we ask for the rewards or not. To fulfil the game requirement, we naturally do not ask for the rewards if $\sum_{s=1}^{t-1} Z_s \geq m$.

THEOREM 23 *Let* $\psi(x) = \exp\left(\frac{\sqrt{m \log K}}{n} x\right)$ *and* $\tilde{g}_{i,t} = \frac{g_{i,t}}{\delta} Z_t$ *with* $\delta = \frac{3m}{4n}$. *Then,*

for any $\varepsilon > 0$, with probability at least $1-\varepsilon$, INF satisfies:

$$R_n \leq 2n\sqrt{\frac{\log K}{m}} + n\sqrt{\frac{27\log(2K\varepsilon^{-1})}{m}},$$

hence

$$\mathbb{E}R_n \leq 8n\sqrt{\frac{\log(6K)}{m}}.$$

This theorem is similar to Theorem 6.2 of [46]. The main difference and novelty is that the policy does not depend on the confidence level, so the high probability bound is valid for any confidence level *for the same policy*, and the expected regret of this policy has also the minimax optimal rate, i.e. $n\sqrt{\frac{\log(K)}{m}}$.

High probability bounds for the bandit game. Here the main difference with Section 4.3 is to use the biased estimates $\tilde{g}_{i,t} = \frac{g_{i,t}}{p_{i,t}}\mathbb{1}_{I_t=i} + \frac{\beta}{p_{i,t}}$ for some appropriate small $\beta > 0$. It may appear surprising as it introduces a bias in the estimate of $g_{i,t}$. However this modification allows to have high probability upper bounds with the correct rate on the difference $\sum_{t=1}^n g_{i,t} - \sum_{t=1}^n \tilde{g}_{i,t}$. A second reason for this modification (but useless for this particular section) is that it allows to track the best expert (see Section 4.3.2). For sake of simplicity, the following theorem concerns deterministic adversaries (which is defined by a fixed matrix of the nK rewards).

THEOREM 24 *For a deterministic adversary, The INF algorithm with $\psi(x) = \left(\frac{3\sqrt{n}}{-x}\right)^2 + \frac{1}{\sqrt{nK}}$ and $\tilde{g}_{i,t} = \frac{g_{i,t}}{p_{i,t}}\mathbb{1}_{I_t=i} + \frac{1}{p_{i,t}\sqrt{nK}}$ satisfies: for any $\varepsilon > 0$, with probability at least $1-\varepsilon$,*

$$R_n \leq 10\sqrt{nK} + 2\sqrt{nK}\log(\varepsilon^{-1}).$$

(Consequently, it also satisfies $\mathbb{E}R_n \leq 12\sqrt{nK}$.)

The novelty of the result, which is similar to Theorem 6.10 of [46], is both the absence of the $\log K$ factor and that the high probability bound is valid for the same policy at any confidence level.

Label efficient and bandit game (LE bandit). In this game first considered by György and Ottucsák [64] and which is a combination of two previously seen games, the forecaster observes the reward of the arm he selected only if he asks for it, and he is not allowed to request it more than m times for some fixed integer number $1 \leq m \leq n$. We consider a similar policy for requesting the reward vector as in the label efficient game. At each round, we draw a Bernoulli random variable Z_t, with parameter $\delta = \frac{3m}{4n}$, to decide whether we ask for the obtained

reward or not. To fulfil the game requirement, we do not ask for the rewards if $\sum_{s=1}^{t-1} Z_s \geq m$.

THEOREM 25 *For* $\psi(x) = \left(\frac{3n}{-\sqrt{m}x}\right)^2 + \frac{1}{\sqrt{nK}}$ *and* $\tilde{g}_{i,t} = g_{i,t}\frac{\mathbb{1}_{I_t=i}}{p_{i,t}}\frac{Z_t}{\delta}$*, the INF algorithm satisfies*

$$\mathbb{E}R_n \leq 40n\sqrt{\frac{K}{m}}.$$

As for the bandit game, the use of the INF forecaster allows to get rid of the $\log K$ factor which was appearing in previous works.

Tracking the best expert in the bandit game. In the previous sections, the cumulative gain of the forecaster was compared to the cumulative gain of the best single expert. Here, it will be compared to more flexible strategies that are allowed to switch actions. A switching strategy is described by a vector $(i_1, \ldots, i_n) \in \{1, \ldots, K\}^n$. Its size is defined by

$$\mathcal{S}(i_1, \ldots, i_n) = \sum_{t=1}^{n-1} \mathbb{1}_{i_{t+1} \neq i_t},$$

and its cumulative gain is

$$G_{(i_1,\ldots,i_n)} = \sum_{t=1}^{n} g_{i_t,t}.$$

The regret of a forecaster with respect to the best switching strategy with S switches is then given by:

$$R_n^S = \max_{(i_1,\ldots,i_n):\, \mathcal{S}(i_1,\ldots,i_n) \leq S} G_{(i_1,\ldots,i_n)} - \sum_{t=1}^{n} g_{I_t,t}.$$

As in Section 4.3.2, we use the estimates

$$\tilde{g}_{i,t} = g_{i,t}\frac{\mathbb{1}_{I_t=i}}{p_{i,t}} + \frac{\beta}{p_{i,t}},$$

and $0 < \beta \leq 1$. The β term, which, as already stated, introduces a bias in the estimate of $g_{i,t}$, constrains the differences $\max_{i=1,\ldots,K} \tilde{G}_{i,t} - \min_{j=1,\ldots,K} \tilde{G}_{j,t}$ to be relatively small. This is the key property in order to track the best switching strategy, provided that the number of switches is not too large. We have the following result for the INF forecaster using the above estimates and an exponential function ψ (recall that for exponential ψ, the INF forecaster reduces to the traditional exponentially weighted forecasters).

THEOREM 26 *Let* $s = S\log\left(\frac{enK}{S}\right) + \log(2K)$ *with* $e = \exp(1)$ *and the natural convention* $S\log(enK/S) = 0$ *for* $S = 0$*. Consider* $\psi(x) = \exp(\eta x) + \frac{\gamma}{K}$ *with* $\gamma = \min\left(\frac{1}{2}, \sqrt{\frac{Ks}{n}}\right)$ *and* $\eta = \sqrt{\frac{s}{20nK}}$*, and the estimates* $\tilde{g}_{i,t} = g_{i,t}\frac{\mathbb{1}_{I_t=i}}{p_{i,t}} + \frac{\beta}{p_{i,t}}$ *with* $\beta = 2\sqrt{\frac{s}{nK}}$*. For these choices, for any* $0 \leq S \leq n-1$*, for any* $\varepsilon > 0$*, with probability at least* $1-\varepsilon$*, INF satisfies:*

$$R_n^S \leq 9\sqrt{nKs} + \sqrt{\frac{nK}{s}}\log(\varepsilon^{-1}),$$

and

$$\mathbb{E}R_n^S \leq 10\sqrt{nKs}.$$

ACKNOWLEDGEMENTS

I am indebted to the Lab directors Renaud Keriven at Ecole des Ponts ParisTech and Jean Ponce at Ecole Normale Suprieure-INRIA for giving me the opportunity to work in an excellent research environment. I am also very grateful to the reviewers of this work: Peter Bartlett, Pascal Massart and Arkadi Nemirovski.

I would like to thank all the colleagues with whom I had the pleasure to work or discuss, in particular, Francis Bach, Sébastien Bubeck, Olivier Catoni, Matthias Hein, Hui Kong, Rmi Munos, Csaba Szepesvári and Sacha Tsybakov. I would also like to thank Yuri Golubev for having accepted to participate and chair the habilitation committee. Special thanks to my loved ones who have constantly and unconditionally supported me while I was developing this research.

Appendix A

Some basic properties of the Kullback-Leibler divergence

The KL divergence between two distributions on some measurable space $\mathcal{G}$

$$K(\rho,\pi) = \begin{cases} \mathbb{E}_{g\sim\rho}\log(\frac{\rho}{\pi}(g)) & \text{if } \rho \ll \pi \\ +\infty & \text{otherwise} \end{cases} \tag{A.1}$$

satisfies for $\rho \ll \pi$, $K(\rho,\pi) = \mathbb{E}_{g\sim\pi}\chi\big(\frac{\rho}{\pi}(g)\big)$, with χ the function defined on $(0,+\infty)$ by $\chi(u) \mapsto u\log(u) + 1 - u$. Since the function χ is nonnegative and equals zero only at 1, we have

$$K(\rho,\pi) \geq 0, \tag{A.2}$$

and

$$K(\rho,\pi) = 0 \Leftrightarrow \rho = \pi. \tag{A.3}$$

Let $h : \mathcal{G} \to \mathbb{R}$ s.t. $\mathbb{E}_{g\sim\pi}e^{h(g)} < +\infty$. Define

$$\pi_h(dg) = \frac{e^{h(g)}}{\mathbb{E}_{g'\sim\pi}e^{h(g')}} \cdot \pi(dg)$$

By expanding the definition of the KL divergence $K(\rho,\pi_h)$, we get

$$K(\rho,\pi_h) = K(\rho,\pi) - \mathbb{E}_{g\sim\rho}h(g) + \log\mathbb{E}_{g\sim\pi}e^{h(g)},$$

which implies from (A.2) and (A.3)

$$\sup_{\rho}\big\{\mathbb{E}_{g\sim\rho}h(g) - K(\rho,\pi)\big\} = \log\mathbb{E}_{g\sim\pi}e^{h(g)}, \tag{A.4}$$

and

$$\text{argmax}_{\rho}\big\{\mathbb{E}_{g\sim\rho}h(g) - K(\rho,\pi)\big\} = \pi_h. \tag{A.5}$$

By differentiating, one may note that the function $\lambda \mapsto K(\pi_{\lambda h},\pi)$ is nondecreasing on $[0,+\infty)$. Finally, if $\mathcal{G}$ is finite and π is the uniform distribution on $\mathcal{G}$, we have

$$K(\rho,\pi) = \log(|\mathcal{G}|) - H(\rho) \leq \log(|\mathcal{G}|), \tag{A.6}$$

where $H(\rho) = -\sum_{g\in\mathcal{G}}\rho(g)\log\rho(g)$ is the Shannon entropy of ρ.

Appendix B

Proof of McAllester's PAC Bayesian bound

McAllester's bound (McA) (p.8) states that with probability at least $1-\varepsilon$, for any $\rho \in \mathcal{M}$, we have

$$\mathbb{E}_{g\sim\rho} R(g) - \mathbb{E}_{g\sim\rho} r(g) \leq \sqrt{\frac{K(\rho,\pi) + \log(2n) + \log(\varepsilon^{-1})}{2n-1}}. \qquad \text{(B.1)}$$

Here is a short proof of this statement that essentially follows the one proposed by Seeger.

Let us first recall that a real-valued random variable V such that $\mathbb{E}e^V \leq 1$ satisfies: for any $\varepsilon > 0$, with probability at least $1-\varepsilon$, we have $V \leq \log(\varepsilon^{-1})$. So to prove (B.1), we only need to check that the random variable

$$V = \sup_{\rho} \Big\{ (2n-1) \big[\max \big(\mathbb{E}_{\rho(df)} R(f) - \mathbb{E}_{\rho(df)} r(f), 0 \big) \big]^2 - K(\rho,\pi) - \log(4n) \Big\}$$

satisfies $\mathbb{E}e^V \leq 1$.

From Jensen's inequality applied to the convex function $x \mapsto \big[\max(x,0)\big]^2$ and the Legendre transform of the KL divergence (A.4), we have

$$\begin{aligned} V &\leq \sup_{\rho} \Big\{ (2n-1) \mathbb{E}_{\rho(df)} [\max \big(R(f) - r(f), 0 \big)]^2 - K(\rho,\pi) - \log(2n) \Big\} \\ &= -\log(2n) + \log \mathbb{E}_{\pi(df)} e^{(2n-1)[\max(R(f)-r(f),0)]^2}, \end{aligned}$$

hence

$$\begin{aligned} \mathbb{E}e^V &\leq \frac{1}{2n} \mathbb{E}\mathbb{E}_{\pi(df)} e^{(2n-1)[\max(R(f)-r(f),0)]^2} \\ &= \frac{1}{2n} \mathbb{E}_{\pi(df)} \Big(1 + \mathbb{E}\big\{ e^{(2n-1)[\max(R(f)-r(f),0)]^2} - 1 \big\} \Big) \qquad \text{from Fubini's theorem} \\ &= \frac{1}{2n} \mathbb{E}_{\pi(df)} \left(1 + \int_0^{+\infty} \mathbb{P}(e^{(2n-1)[\max(R(f)-r(f),0)]^2} - 1 > t) dt \right) \\ &= \frac{1}{2n} \mathbb{E}_{\pi(df)} \left(1 + \int_0^{+\infty} \mathbb{P}\left(R(f) - r(f) > \sqrt{\frac{\log(t+1)}{2n-1}} \right) dt \right) \\ &\leq \frac{1}{2n} \mathbb{E}_{\pi(df)} \left(1 + \int_0^{+\infty} e^{-2n\frac{\log(t+1)}{2n-1}} dt \right) \qquad \text{from Hoeffding's inequality} \\ &= \frac{1}{2n} \mathbb{E}_{\pi(df)} \left(1 + \int_0^{+\infty} (t+1)^{-\frac{2n}{2n-1}} dt \right) \\ &= 1, \end{aligned}$$

which ends the proof.

Appendix C

Proof of Seeger's PAC Bayesian bound

Here we sketch the proof of (S) (p.8), which states that with probability at least $1-\varepsilon$, for any $\rho \in \mathcal{M}$, we have

$$K(\mathbb{E}_{g\sim\rho} r(g) || \mathbb{E}_{g\sim\rho} R(g)) \leq \frac{K(\rho,\pi) + \log(2\sqrt{n}\varepsilon^{-1})}{n}, \tag{C.1}$$

where $K(q||p) = K(\mathrm{Be}(q), \mathrm{Be}(p))$ with $\mathrm{Be}(q)$ and $\mathrm{Be}(p)$ denoting the Bernoulli distributions of parameter q and p. The proof follows the same line as the one of (McA). We introduce

$$V = \sup_{\rho} \Big\{ nK(\mathbb{E}_{\rho(df)} r(f)) || \mathbb{E}_{\rho(df)} R(f)) - K(\rho,\pi) - \log(2\sqrt{n}) \Big\},$$

and as in the previous proof, we only need to check that $\mathbb{E}e^V \leq 1$. This is done by using Jensen's inequality for the convex function $(q,p) \mapsto K(q||p)$ and using the Legendre transform of the KL divergence (A.4). We have

$$\begin{aligned}
\mathbb{E}e^V &\leq \mathbb{E}e^{\sup_\rho \left\{ n\mathbb{E}_{\rho(df)} K(r(f)||R(f)) - K(\rho,\pi) - \log(2\sqrt{n}) \right\}} \\
&= \frac{1}{2\sqrt{n}} \mathbb{E}\mathbb{E}_{\pi(df)} e^{nK(r(f))||R(f))} \\
&= \frac{1}{2\sqrt{n}} \mathbb{E}_{\pi(df)} \sum_{k=0}^{n} \mathbb{P}(nr(f)=k) \Big(\frac{k}{nR(f)} \Big)^k \Big(\frac{n-k}{n[1-R(f)]} \Big)^{n-k} \\
&= \frac{1}{2\sqrt{n}} \mathbb{E}_{\pi(df)} \sum_{k=0}^{n} \binom{n}{k} \Big(\frac{k}{n} \Big)^k \Big(\frac{n-k}{n} \Big)^{n-k} \\
&\leq 1,
\end{aligned}$$

where the last inequality is obtained from computations using Stirling's approximation.

The same procedure can be used to prove the other PAC-Bayesian bounds of Chapter 2, Section 2.2. A similar way of approaching PAC-Bayesian theorems is given in [60].

Appendix D

Proof of the learning rate of the progressive mixture rule

Here is the proof in a concise form under the boundedness assumptions of Theorem 6 that the expected excess risk of the progressive mixture rule is upper bounded by $\frac{\log d}{\lambda(n+1)}$ for $\lambda \geq \frac{1}{8}$. The condition on λ guarantees that for any $y \in [-1,1]$, the function $y' \mapsto e^{-\lambda(y-y')^2}$ is concave on $[-1,1]$. Thus we can write

$$
\begin{aligned}
&\mathbb{E}R\bigg(\frac{1}{n+1}\sum_{i=0}^{n}\mathbb{E}_{g\sim\pi_{-\lambda\Sigma_i}}\, g\bigg)\\
\leq&\frac{1}{n+1}\sum_{i=0}^{n}\mathbb{E}R\big(\mathbb{E}_{g\sim\pi_{-\lambda\Sigma_i}}\, g\big) &\text{(D.1)}\\
=&\frac{1}{n+1}\sum_{i=0}^{n}\mathbb{E}_{Z_1^{i+1}}[Y_{i+1}-\mathbb{E}_{g\sim\pi_{-\lambda\Sigma_i}}\, g(X_{i+1})]^2 &\text{(D.2)}\\
=&\frac{1}{n+1}\mathbb{E}_{Z_1^{n+1}}\sum_{i=0}^{n}[Y_{i+1}-\mathbb{E}_{g\sim\pi_{-\lambda\Sigma_i}}\, g(X_{i+1})]^2 &\text{(D.3)}\\
\leq&\frac{1}{n+1}\mathbb{E}_{Z_1^{n+1}}\sum_{i=0}^{n}\bigg\{-\frac{1}{\lambda}\log\mathbb{E}_{g\sim\pi_{-\lambda\Sigma_i}}\, e^{-\lambda[Y_{i+1}-g(X_{i+1})]^2}\bigg\} &\text{(D.4)}\\
=&\frac{1}{\lambda(n+1)}\mathbb{E}_{Z_1^{n+1}}\sum_{i=0}^{n}\log\left(\frac{\mathbb{E}_{g\sim\pi}\, e^{-\lambda\Sigma_i(g)}}{\mathbb{E}_{g\sim\pi}\, e^{-\lambda\Sigma_{i+1}(g)}}\right)\\
=&-\frac{1}{\lambda(n+1)}\mathbb{E}_{Z_1^{n+1}}\log\mathbb{E}_{g\sim\pi}\, e^{-\lambda\Sigma_{n+1}(g)} &\text{(D.5)}\\
\leq&-\frac{1}{\lambda(n+1)}\mathbb{E}_{Z_1^{n+1}}\log\left(\frac{e^{-\lambda\Sigma_{n+1}(g^*_{\mathbf{MS}})}}{d}\right)\\
=&R(g^*_{\mathbf{MS}})+\frac{\log d}{\lambda(n+1)},
\end{aligned}
$$

where (D.1) comes from Jensen's inequality on the convex function $y' \mapsto (y-y')^2$, (D.2) uses that the distribution $\pi_{-\lambda\Sigma_i}$ depends only on Z_1^i, (D.4) comes from Jensen's inequality on the concave function $y' \mapsto e^{-\lambda(y-y')^2}$, and (D.5) is the core of the proof and explains why PM is based on a Cesaro mean. The steps (D.2) and (D.3) are exactly the two steps of the proof of Lemma 7. Note that this analysis gives a result similar to the one in Theorem 6, except that the factor 2 is replaced by $\frac{1}{\lambda} \geq 8$. For the progressive indirect mixture rule, $\mathbb{E}_{g\sim\pi_{-\lambda\Sigma_i}}$ are replaced by $\hat{h}_i$, and the step (D.4) is still valid from the very definition (3.2.1) of $\hat{h}_i$.

Appendix E

The empirical Bernstein's inequality

The goal of the empirical Bernstein's inequality is to provide confidence bounds on the expectation of a distribution with bounded support, say in $[0,1]$, given a sample from it. Let $U, U_1, U_2, \dots$ be independent and identically distributed random variables taking their values in $[0,1]$. Let

$$\bar{U}_t = \frac{1}{t}\sum_{i=1}^{t} U_i \qquad \text{and} \qquad \bar{V}_t = \frac{1}{t}\sum_{i=1}^{t}(U_i - \bar{U}_t)^2.$$

Here we prove the empirical Bernstein's inequality (Lemma 21, p.51), which states that for any $\varepsilon > 0$, with probability at least $1 - 2\varepsilon$, for any $t \in \{1, \dots, n\}$ and $\bar{\ell}_t = \frac{n \log(\varepsilon^{-1})}{t^2}$, we have

$$\bar{U}_t - \mathbb{E}U < \min\left(\sqrt{2\bar{\ell}_t(\bar{V}_t + \bar{\ell}_t)} + \bar{\ell}_t\Big(\frac{1}{3} + \sqrt{1 - 3\bar{V}_t}\Big)\ ,\ \sqrt{\frac{\bar{\ell}_t}{2}} \right). \tag{E.1}$$

PROOF. Let $\Lambda(\lambda) = \log \mathbb{E}e^{\lambda(U - \mathbb{E}U)}$ be the log-Laplace transform of the random variable $U - \mathbb{E}U$. Let $S_t = \sum_{i=1}^{t}(U_i - \mathbb{E}U_i)$ with the convention $S_0 = 0$. From Inequality (2.17) of [67], we have[1]

$$\mathbb{P}\big(\max_{1 \le t \le n} S_t \ge s\big) \le \inf_{\lambda > 0} e^{-\lambda s + n\Lambda(\lambda)}.$$

Let $V = \mathbb{V}\mathrm{ar}\, U$. Hoeffding's inequality and Bennett's inequality implies

$$\Lambda(\lambda) \le \min\Big(\frac{\lambda^2}{8}, (e^\lambda - 1 - \lambda)V\Big),$$

which by standard computations (see, e.g., Inequality (45) of [15]) gives that for any $\varepsilon > 0$, with probability at least $1 - \varepsilon$,

$$\max_{1 \le t \le n} S_t < \min\left(\sqrt{\frac{n \log(\varepsilon^{-1})}{2}}, \sqrt{2nV \log(\varepsilon^{-1})} + \frac{\log(\varepsilon^{-1})}{3} \right). \tag{E.2}$$

[1] This comes from a martingale argument due to Doob. For any $\lambda > 0$, the sequence $(e^{\lambda S_t - t\Lambda(\lambda)})_{t \ge 0}$ is a martingale with respect to the filtration $\big(\sigma(U_1, \dots, U_t)\big)_{t \ge 0}$ since $\mathbb{E}(e^{\lambda S_t - t\Lambda(\lambda)} | U_1, \dots, U_{t-1}) = e^{\lambda S_{t-1} - (t-1)\Lambda(\lambda)}$. Introduce the stopping time $T = \min\big(n + 1, \min\{t \in \mathbb{N} : S_t \ge s\}\big)$. From the optional stopping theorem, for any $\lambda > 0$, we have

$$1 = \mathbb{E}e^{\lambda S_T - T\Lambda(\lambda)} \ge \mathbb{P}(T \le n) e^{\lambda s - n\Lambda(\lambda)},$$

hence

$$\mathbb{P}\big(\max_{1 \le t \le n} S_t \ge s\big) = \mathbb{P}(T \le n) \le \inf_{\lambda > 0} e^{-\lambda s + n\Lambda(\lambda)}.$$

Let $W = (U - \mathbb{E}U)^2$ and $W_i = (U_i - \mathbb{E}U_i)^2$ for $i \geq 1$. Let $S'_t = \sum_{i=1}^t (-W_i + \mathbb{E}W_i)$ and $\Lambda'(\lambda) = \log \mathbb{E}e^{\lambda(-W+\mathbb{E}W)}$. As above, from Inequality (2.17) of [67], we have

$$\mathbb{P}\big(\max_{1\leq t\leq n} S'_t \geq s\big) \leq \inf_{\lambda>0} e^{-\lambda s + n\Lambda'(\lambda)}.$$

Now using that $e^{-u} \leq 1 - u + \frac{u^2}{2}$ for $u \geq 0$ and $\log(1+u) \leq u$ from $u \geq -1$, we have $\log \mathbb{E}e^{-\lambda W} \leq \log \mathbb{E}(1 - \lambda W + \frac{\lambda^2 W^2}{2}) \leq -\lambda \mathbb{E}W + \frac{\lambda^2}{2}\mathbb{E}(W^2)$, hence $\Lambda'(\lambda) \leq \frac{\lambda^2}{2}\mathbb{E}(W^2)$. Optimizing with respect to λ gives that for any $\varepsilon > 0$, with probability at least $1 - \varepsilon$,

$$\max_{1\leq t\leq n} S'_t < \sqrt{2n\mathbb{E}(W^2)\log(\varepsilon^{-1})}. \tag{E.3}$$

Now we use the following lemma to bound $\mathbb{E}(W^2)$.

LEMMA 27 *A random variable U taking its values in $[0,1]$ satisfies*

$$\mathbb{E}[(U - \mathbb{E}U)^4] \leq V(1 - 3V), \tag{E.4}$$

where $V = \mathbb{E}[(U - \mathbb{E}U)^2]$ is the variance of U. If U admits a Bernoulli distribution, one can put an equality in (E.4)*.*

PROOF. We have

$$\begin{aligned}\mathbb{E}[(U - \mathbb{E}U)^4] - V(1-3V) = \mathbb{E}\big([U^3 - U + \mathbb{E}(U)][U - \mathbb{E}(U)]\big) \\ + 3\big([\mathbb{E}(U^2)]^2 - \mathbb{E}(U)\mathbb{E}(U^3)\big).\end{aligned}$$

From Chebyshev's association inequality (also referred to as the Fortuin-Kasteleyn-Ginibre inequality), both terms in the right-hand side are nonpositive. An alternative proof consists in expanding the terms in Lemma 8 of [101] and noticing that this exactly gives (E.4). The result for Bernoulli distributions comes from direct computations. □

Combining the above lemma with (E.3), we get that with probability at least $1 - \varepsilon$,

$$\max_{1\leq t\leq n} S'_t < \sqrt{2nV(1-3V)\log(\varepsilon^{-1})}. \tag{E.5}$$

We now work on the event $\mathcal{E}$ of probability at least $1 - 2\varepsilon$ on which both (E.5) and (E.2) hold. The variance decomposition gives $\bar{V}_t = \frac{1}{t}\sum_{i=1}^t (U_i - \bar{U}_t)^2 = -(\mathbb{E}U - \bar{U}_t)^2 + \frac{1}{t}\sum_{i=1}^t W_i$, hence $S'_t = t(V - \bar{V}_t) - t(\mathbb{E}U - \bar{U}_t)^2$. For any $1 \leq t \leq n$, we have

$$\bar{U}_t - \mathbb{E}U < \min\left(\sqrt{\frac{\bar{\ell}_t}{2}}, \sqrt{2V\bar{\ell}_t} + \frac{\bar{\ell}_t}{3}\right), \tag{E.6}$$

and

$$V - \bar{V}_t < \sqrt{2V(1-3V)\bar{\ell}_t} + (\bar{U}_t - \mathbb{E}U)^2 \tag{E.7}$$

If $\bar{U}_t < \mathbb{E}U$, then (E.1) is trivial. If $\bar{V}_t \geq V$, (E.1) is a direct consequence of (E.6) (since $\frac{4}{3} - 3\bar{V}_t \geq \frac{4}{3} - \frac{3}{4} > \frac{1}{3}$). Therefore, from now and on, we consider $\bar{U}_t \geq \mathbb{E}U$ and $\bar{V}_t < V$. Then (E.6) implies $(\bar{U}_t - \mathbb{E}U)^2 \leq \bar{\ell}_t/2$, and (E.7) leads to

$$\bar{V}_t > V - \sqrt{2V(1-3\bar{V}_t)\bar{\ell}_t} - \bar{\ell}_t/2 = \left(\sqrt{V} - \sqrt{\frac{\bar{\ell}_t(1-3\bar{V}_t)}{2}}\right)^2 - \frac{\bar{\ell}_t(2-3\bar{V}_t)}{2},$$

hence

$$\sqrt{V} < \sqrt{\bar{V}_t + \frac{\bar{\ell}_t(2-3\bar{V}_t)}{2}} + \sqrt{\frac{\bar{\ell}_t(1-3\bar{V}_t)}{2}} \leq \sqrt{\bar{V}_t + \bar{\ell}_t} + \sqrt{\frac{\bar{\ell}_t(1-3\bar{V}_t)}{2}}.$$

By plugging this inequality into (E.6), we get (E.2). For the two-sided inequality (4.2.12), one just needs to add the same inequality as (E.6) for $-\bar{U}_t$. At the end, three maximal inequalities are used (corresponding to U_i, $-U_i$ and $-W_i$), so that the result holding with probability at least $1-\varepsilon$ contains $\log(3\varepsilon^{-1})$ terms. $\square$

Appendix F

On Exploration-Exploitation with Exponential weights (EXP3)

F.1. THE VARIANTS OF EXP3

Parameters: $\eta \in (0, 1/K]$ and $\gamma \in [0, 1]$.
Let p_1 be the uniform distribution over $\{1, \ldots, K\}$.

For each round $t = 1, 2, \ldots,$

(1) Draw an arm I_t according to the probability distribution p_t.

(2) Compute the estimated gain for each arm:

$$\tilde{g}_{i,t} = \begin{cases} \frac{g_{i,t}}{p_{i,t}} \mathbb{1}_{I_t=i} & \text{for the reward-magnifying version of EXP3} \\ 1 - \frac{1-g_{i,t}}{p_{i,t}} \mathbb{1}_{I_t=i} & \text{for the loss-magnifying version of EXP3} \\ \frac{g_{i,t}}{p_{i,t}} \mathbb{1}_{I_t=i} + \frac{\beta}{p_{i,t}} & \text{for the tracking version of EXP3} \\ \frac{g_{i,t}(1+\beta \frac{g_{i,t}}{p_{i,t}})}{p_{i,t}} \mathbb{1}_{I_t=i} & \text{for the tightly biased version of EXP3} \end{cases}$$

and update the estimated cumulative gain: $\tilde{G}_{i,t} = \sum_{s=1}^{t} \tilde{g}_{i,s}$.

(3) Compute the new probability distribution over the arms:

$$p_{t+1} = \gamma p_1 + (1 - \gamma) q_{t+1},$$

with

$$q_{i,t+1} = \frac{\exp\left(\eta \tilde{G}_{i,t}\right)}{\sum_{k=1}^{K} \exp\left(\eta \tilde{G}_{k,t}\right)}.$$

Figure F.1: EXP3 (Exploration-Exploitation with Exponential weights) for the adversarial bandit problem.

There are several variants of EXP3. They differ by the way $g_{i,t}$ is estimated as shown in Figure F.1. For deterministic adversaries, the loss-magnifying version of EXP3 has the advantage to provide the best known constant in front of the $\sqrt{nK \log K}$ term, that is $\sqrt{2}$ (note that our work succeeds in removing the $\log K$ term but at the price of a larger numerical constant factor). For deterministic adversaries, the reward-magnifying version of EXP3 (which is the one in the

seminal paper of Auer, Cesa-Bianchi, Freund and Schapire [19] for $\gamma = K\eta$) has the advantage that the factor n in $\sqrt{nK\log K}$ can be replaced by $\max_{i=1,\dots,n} G_{i,n}$, where $G_{i,n} = \sum_{t=1}^{n} g_{i,t}$. The tracking version of EXP3 is the one proposed in Section 6.8 of [46] (and the one presented in Section 4.3.2). It (slightly) overestimates the rewards since we have $\mathbb{E}_{I_t\sim p_t}\tilde{g}_{i,t} = g_{i,t} + \frac{\beta}{p_{i,t}}$. This idea was introduced in [17] for tracking the best expert. In [12], we have introduced the tightly biased version of EXP3 to achieve regret bounds depending on the performance of the optimal arm. Contrarily to the reward-magnifying version of EXP3, these bounds hold for any adversary and high probability regret bounds are also obtained.

F.2. Proof of the learning rate of the reward-magnifying EXP3

Here we give an analysis of the reward-magnifying EXP3 (defined in Figure F.1), which is an improvement (in terms of constant only) of the one in [19, Section 3].

THEOREM 28 *Let $G_{\max} = \max_{i=1,\dots,K} G_{i,n}$. For deterministic adversaries, if $4\eta K \leq 5\gamma$, the expected regret of the reward-magnifying EXP3 satisfies*

$$\mathbb{E}R_n \leq \frac{\log K}{\eta} + \gamma G_{\max}.$$

In particular, if $\eta = \sqrt{\frac{5\log K}{4nK}}$ and $\gamma = \min\left(\sqrt{\frac{4K\log K}{5n}}, 1\right)$, we have

$$\mathbb{E}R_n \leq \sqrt{\frac{16}{5}nK\log(K)}.$$

PROOF. The condition $4\eta K \leq 5\gamma$ is put to guarantee that $\Psi\left(\frac{\eta K}{\gamma}\right) \leq \frac{\gamma}{\eta K}$, where $\Psi : u \mapsto \frac{e^u - 1 - u}{u^2}$ is an increasing function. For any adversary, we have

$$\begin{aligned}
\sum_{t=1}^{n} g_{I_t,t} &= \sum_{t=1}^{n} \mathbb{E}_{k\sim p_t}\tilde{g}_{k,t} \\
&= \frac{1-\gamma}{\eta}\sum_{t=1}^{n}\left(\log \mathbb{E}_{i\sim q_t} e^{\eta\tilde{g}_{i,t}} - \log\left[e^{-\frac{\eta}{1-\gamma}\mathbb{E}_{k\sim p_t}\tilde{g}_{k,t}}\mathbb{E}_{i\sim q_t} e^{\eta\tilde{g}_{i,t}}\right]\right) \\
&= \frac{1-\gamma}{\eta}\left(S - \sum_{t=1}^{n}\log(D_t)\right),
\end{aligned}$$

where

$$S = \sum_{t=1}^{n}\log \mathbb{E}_{i\sim q_t} e^{\eta\tilde{g}_{i,t}} = \sum_{t=1}^{n}\log\left(\frac{\mathbb{E}_{i\sim p_1} e^{\eta\tilde{G}_{i,t}}}{\mathbb{E}_{i\sim p_1} e^{\eta\tilde{G}_{i,t-1}}}\right) = \log \mathbb{E}_{i\sim p_1} e^{\eta\tilde{G}_{i,n}}$$

and

$$
\begin{aligned}
D_t &= e^{-\frac{\eta}{1-\gamma}\mathbb{E}_{k\sim p_t}\tilde{g}_{k,t}}\mathbb{E}_{i\sim q_t}e^{\eta\tilde{g}_{i,t}} \\
&\leq e^{-\frac{\eta}{1-\gamma}\mathbb{E}_{k\sim p_t}\tilde{g}_{k,t}}\mathbb{E}_{i\sim q_t}\left(1+\eta\tilde{g}_{i,t}+\Psi\Big(\frac{\eta K}{\gamma}\Big)\eta^2\tilde{g}_{i,t}^2\right) && \text{(F.1)} \\
&= e^{-\frac{\eta}{1-\gamma}\mathbb{E}_{k\sim p_t}\tilde{g}_{k,t}}\left(1+\eta\frac{\mathbb{E}_{i\sim p_t}\tilde{g}_{i,t}-\gamma\mathbb{E}_{i\sim p_1}\tilde{g}_{i,t}}{1-\gamma}+\Psi\Big(\frac{\eta K}{\gamma}\Big)\eta^2\mathbb{E}_{i\sim q_t}\tilde{g}_{i,t}^2\right) \\
&\leq e^{-\frac{\eta}{1-\gamma}\mathbb{E}_{k\sim p_t}\tilde{g}_{k,t}}\left(1+\frac{\eta}{1-\gamma}\mathbb{E}_{i\sim p_t}\tilde{g}_{i,t}-\frac{\eta\gamma}{1-\gamma}\mathbb{E}_{i\sim p_1}\tilde{g}_{i,t}+\frac{\Psi(\frac{\eta K}{\gamma})\eta^2 K}{1-\gamma}\mathbb{E}_{i\sim p_1}\tilde{g}_{i,t}\right) && \text{(F.2)} \\
&\leq e^{-\frac{\eta}{1-\gamma}\mathbb{E}_{k\sim p_t}\tilde{g}_{k,t}}\left(1+\frac{\eta}{1-\gamma}\mathbb{E}_{i\sim p_t}\tilde{g}_{i,t}\right) && \text{(F.3)} \\
&\leq 1.
\end{aligned}
$$

To get (F.1), we used that Ψ is an increasing function and that $\eta\tilde{g}_{i,t} \leq \frac{\eta}{p_{i,t}} \leq \frac{\eta K}{\gamma}$. For (F.2), we used $(1-\gamma)\mathbb{E}_{i\sim q_t}\tilde{g}_{i,t}^2 \leq \mathbb{E}_{i\sim p_t}\tilde{g}_{i,t}^2 = \frac{g_{I_t,t}^2}{p_{I_t,t}} \leq \sum_{i=1}^K \tilde{g}_{i,t} = K\mathbb{E}_{i\sim p_1}\tilde{g}_{i,t}$. For (F.3), we used $\eta K\Psi\big(\frac{\eta K}{\gamma}\big) \leq \gamma$. We have thus proved

$$
\sum_{t=1}^n g_{I_t,t} \geq \frac{1-\gamma}{\eta}\log\mathbb{E}_{i\sim p_1}e^{\eta\tilde{G}_{i,n}} \qquad \text{(F.4)}
$$

For a deterministic adversary, we have $\mathbb{E}\tilde{G}_{i,n} = \mathbb{E}G_{i,n} = G_{i,n}$, so that

$$
\begin{aligned}
\mathbb{E}\sum_{t=1}^n g_{I_t,t} &\geq \frac{1-\gamma}{\eta}\mathbb{E}\log\mathbb{E}_{i\sim p_1}e^{\eta\tilde{G}_{i,n}} \\
&\geq \frac{1-\gamma}{\eta}\log\mathbb{E}_{i\sim p_1}e^{\eta\mathbb{E}\tilde{G}_{i,n}} && \text{(F.5)} \\
&= \frac{1-\gamma}{\eta}\log\mathbb{E}_{i\sim p_1}e^{\eta G_{i,n}} \geq -\frac{(1-\gamma)\log K}{\eta} + (1-\gamma)\max_{i=1,\ldots,K} G_{i,n},
\end{aligned}
$$

where Inequality (F.5) which moves the expectation sign inside the exponential can be viewed as an infinite dimensional Jensen's inequality (see Lemma 3.2 of [9]). For a deterministic adversary, we have proved

$$
\mathbb{E}R_n = \max_{i=1,\ldots,K} G_{i,n} - \mathbb{E}\sum_{t=1}^n g_{I_t,t} \leq \frac{(1-\gamma)\log K}{\eta} + \gamma\max_{i=1,\ldots,K} G_{i,n},
$$

hence the first claimed result.

The second result is trivial when $\sqrt{4K\log K/(5n)} \geq 1$ since the upper bound is then larger than n. Otherwise, we have $\gamma = \sqrt{4K\log K/(5n)} < 1$ and $4\eta K = 5\gamma$ so that the result follows from the first one. $\square$

Appendix G

Experimental results for the min-max truncated estimator defined in Section 3.4.2

In Section G.1, we detail the different kinds of noises we work with. Then, Sections G.2, G.3 and G.4 describe the three types of functional relationships between the input, the output and the noise involved in our experiments. A motivation for choosing these input-output distributions was the ability to compute exactly the excess risk, and thus to compare easily estimators. Section G.5 presents the experimental results.

G.1. NOISE DISTRIBUTIONS

In our experiments, we consider different types of noise that are centered and with unit variance:

- the standard Gaussian noise: $W \sim \mathcal{N}(0,1)$,
- a heavy-tailed noise defined by: $W = \mathrm{sign}(V)/|V|^{1/q}$, with $V \sim \mathcal{N}(0,1)$ a standard Gaussian random variable and $q = 2.01$ (the real number q is taken strictly larger than 2 as for $q = 2$, the random variable W would not admit a finite second moment).
- an asymmetric heavy-tailed noise defined by:

$$W = \begin{cases} |V|^{-1/q} & \text{if } V > 0, \\ -\frac{q}{q-1} & \text{otherwise,} \end{cases}$$

  with $q = 2.01$ with $V \sim \mathcal{N}(0,1)$ a standard Gaussian random variable.
- a mixture of a Dirac random variable with a low-variance Gaussian random variable defined by: with probability p, $W = \sqrt{(1-\rho)/p}$, and with probability $1-p$, W is drawn from

$$\mathbb{N}\left(-\frac{\sqrt{p(1-\rho)}}{1-p}, \frac{\rho}{1-p} - \frac{p(1-\rho)}{(1-p)^2} \right).$$

  The parameter $\rho \in [p,1]$ characterizes the part of the variance of W explained by the Gaussian part of the mixture. Note that this noise admits exponential moments, but for n of order $1/p$, the Dirac part of the mixture generates low signal to noise points.

G.2. INDEPENDENT NORMALIZED COVARIATES (INC(n, d))

In INC(n, d), the input-output pair is such that

$$Y = \langle \theta^*, X \rangle + \sigma W,$$

where the components of X are independent standard normal distributions, $\theta^* = (10, \dots, 10)^T \in \mathbb{R}^d$, and $\sigma = 10$.

G.3. HIGHLY CORRELATED COVARIATES (HCC(n, d))

In HCC(n, d), the input-output pair is such that

$$Y = \langle \theta^*, X \rangle + \sigma W,$$

where X is a multivariate centered normal Gaussian with covariance matrix Q obtained by drawing a (d, d)-matrix A of uniform random variables in $[0, 1]$ and by computing $Q = AA^T$, $\theta^* = (10, \dots, 10)^T \in \mathbb{R}^d$, and $\sigma = 10$. So the only difference with the setting of Section G.2 is the correlation between the covariates.

G.4. TRIGONOMETRIC SERIES (TS(n, d))

Let X be a uniform random variable on $[0, 1]$. Let d be an even number. Let

$$\vec{g}(X) = \big(\cos(2\pi X), \dots, \cos(d\pi X), \sin(2\pi X), \dots, \sin(d\pi X) \big)^T.$$

In TS(n, d), the input-output pair is such that

$$Y = 20X^2 - 10X - \frac{5}{3} + \sigma W,$$

with $\sigma = 10$. One can check that this implies

$$\theta^* = \left(\frac{20}{\pi^2}, \dots, \frac{20}{\pi^2 (\frac{d}{2})^2}, -\frac{10}{\pi}, \dots, -\frac{10}{\pi(\frac{d}{2})} \right)^T \in \mathbb{R}^d.$$

G.5. RESULTS

Tables G.1 and G.2 give the results for the mixture noise. Tables G.3, G.4 and G.5 provide the results for the heavy-tailed noise and the standard Gaussian noise. Each line of the tables has been obtained after 1000 generations of the training

set. These results show that the min-max truncated estimator $\hat{g}$ is often equal to the ordinary least squares estimator $\hat{g}^{(\text{ols})}$, while it ensures impressive consistent improvements when it differs from $\hat{g}^{(\text{ols})}$. In this latter case, the number of points that are not considered in $\hat{g}$, i.e. the number of points with low signal to noise ratio, varies a lot from 1 to 150 and is often of order 30. Note that not only the points that we expect to be considered as outliers (i.e. very large output points) are erased, and that these points seem to be taken out by local groups: see Figures G.1 and G.2 in which the erased points are marked by surrounding circles.

Besides, the heavier the noise tail is (and also the larger the variance of the noise is), the more often the truncation modifies the initial ordinary least squares estimator, and the more improvements we get from the min-max truncated estimator, which also becomes much more robust than the ordinary least squares estimator (see the confidence intervals in the tables).

Table G.1: Comparison of the min-max truncated estimator $\hat{g}$ with the ordinary least squares estimator $\hat{g}^{(\text{ols})}$ for the mixture noise (see Section G.1) with $\rho = 0.1$ and $p = 0.005$. In parenthesis, the 95%-confidence intervals for the estimated quantities.

	nb of iterations	nb of iter. with $R(\hat{g}) \neq R(\hat{g}^{(\text{ols})})$	nb of iter. with $R(\hat{g}) < R(\hat{g}^{(\text{ols})})$	$\mathbb{E}R(\hat{g}^{(\text{ols})}) - R(g^*_{\mathbf{L}})$	$\mathbb{E}R(\hat{g}) - R(g^*_{\mathbf{L}})$	$\mathbb{E}R[(\hat{g}^{(\text{ols})}) \vert \hat{g} \neq \hat{g}^{(\text{ols})}] - R(g^*_{\mathbf{L}})$	$\mathbb{E}[R(\hat{g}) \vert \hat{g} \neq \hat{g}^{(\text{ols})}] - R(g^*_{\mathbf{L}})$
INC(n=200,d=1)	1000	419	405	$0.567(\pm 0.083)$	$0.178(\pm 0.025)$	$1.191(\pm 0.178)$	$0.262(\pm 0.052)$
INC(n=200,d=2)	1000	506	498	$1.055(\pm 0.112)$	$0.271(\pm 0.030)$	$1.884(\pm 0.193)$	$0.334(\pm 0.050)$
HCC(n=200,d=2)	1000	502	494	$1.045(\pm 0.103)$	$0.267(\pm 0.024)$	$1.866(\pm 0.174)$	$0.316(\pm 0.032)$
TS(n=200,d=2)	1000	561	554	$1.069(\pm 0.089)$	$0.310(\pm 0.027)$	$1.720(\pm 0.132)$	$0.367(\pm 0.036)$
INC(n=1000,d=2)	1000	402	392	$0.204(\pm 0.015)$	$0.109(\pm 0.008)$	$0.316(\pm 0.029)$	$0.081(\pm 0.011)$
INC(n=1000,d=10)	1000	950	946	$1.030(\pm 0.041)$	$0.228(\pm 0.016)$	$1.051(\pm 0.042)$	$0.207(\pm 0.014)$
HCC(n=1000,d=10)	1000	942	942	$0.980(\pm 0.038)$	$0.222(\pm 0.015)$	$1.008(\pm 0.039)$	$0.203(\pm 0.015)$
TS(n=1000,d=10)	1000	976	973	$1.009(\pm 0.037)$	$0.228(\pm 0.017)$	$1.018(\pm 0.038)$	$0.217(\pm 0.016)$
INC(n=2000,d=2)	1000	209	207	$0.104(\pm 0.007)$	$0.078(\pm 0.005)$	$0.206(\pm 0.021)$	$0.082(\pm 0.012)$
HCC(n=2000,d=2)	1000	184	183	$0.099(\pm 0.007)$	$0.076(\pm 0.005)$	$0.196(\pm 0.023)$	$0.070(\pm 0.010)$
TS(n=2000,d=2)	1000	172	171	$0.101(\pm 0.007)$	$0.080(\pm 0.005)$	$0.206(\pm 0.020)$	$0.083(\pm 0.012)$
INC(n=2000,d=10)	1000	669	669	$0.510(\pm 0.018)$	$0.206(\pm 0.012)$	$0.572(\pm 0.023)$	$0.117(\pm 0.009)$
HCC(n=2000,d=10)	1000	669	669	$0.499(\pm 0.018)$	$0.207(\pm 0.013)$	$0.561(\pm 0.023)$	$0.125(\pm 0.011)$
TS(n=2000,d=10)	1000	754	753	$0.516(\pm 0.018)$	$0.195(\pm 0.013)$	$0.558(\pm 0.022)$	$0.131(\pm 0.011)$

Table G.2: Comparison of the min-max truncated estimator $\hat{g}$ with the ordinary least squares estimator $\hat{g}^{(\text{ols})}$ for the mixture noise (see Section G.1) with $\rho = 0.4$ and $p = 0.005$. In parenthesis, the 95%-confidence intervals for the estimated quantities.

	nb of iterations	nb of iter. with $R(\hat{g}) \neq R(\hat{g}^{(\text{ols})})$	nb of iter. with $R(\hat{g}) < R(\hat{g}^{(\text{ols})})$	$\mathbb{E}R(\hat{g}^{(\text{ols})}) - R(g_{\mathbf{L}}^*)$	$\mathbb{E}R(\hat{g}) - R(g_{\mathbf{L}}^*)$	$\mathbb{E}R[(\hat{g}^{(\text{ols})})\|\hat{g} \neq \hat{g}^{(\text{ols})}] - R(g_{\mathbf{L}}^*)$	$\mathbb{E}[R(\hat{g})\|\hat{g} \neq \hat{g}^{(\text{ols})}] - R(g_{\mathbf{L}}^*)$
INC(n=200,d=1)	1000	234	211	$0.551(\pm 0.063)$	$0.409(\pm 0.042)$	$1.211(\pm 0.210)$	$0.606(\pm 0.110)$
INC(n=200,d=2)	1000	195	186	$1.046(\pm 0.088)$	$0.788(\pm 0.061)$	$2.174(\pm 0.293)$	$0.848(\pm 0.118)$
HCC(n=200,d=2)	1000	222	215	$1.028(\pm 0.079)$	$0.748(\pm 0.051)$	$2.157(\pm 0.243)$	$0.897(\pm 0.112)$
TS(n=200,d=2)	1000	291	268	$1.053(\pm 0.079)$	$0.805(\pm 0.058)$	$1.701(\pm 0.186)$	$0.851(\pm 0.093)$
INC(n=1000,d=2)	1000	127	117	$0.201(\pm 0.013)$	$0.181(\pm 0.012)$	$0.366(\pm 0.053)$	$0.207(\pm 0.035)$
INC(n=1000,d=10)	1000	262	249	$1.023(\pm 0.035)$	$0.902(\pm 0.030)$	$1.238(\pm 0.081)$	$0.777(\pm 0.054)$
HCC(n=1000,d=10)	1000	201	192	$0.991(\pm 0.033)$	$0.902(\pm 0.031)$	$1.235(\pm 0.088)$	$0.790(\pm 0.067)$
TS(n=1000,d=10)	1000	171	162	$1.009(\pm 0.033)$	$0.951(\pm 0.031)$	$1.166(\pm 0.098)$	$0.825(\pm 0.071)$
INC(n=2000,d=2)	1000	80	77	$0.105(\pm 0.007)$	$0.099(\pm 0.006)$	$0.214(\pm 0.042)$	$0.135(\pm 0.029)$
HCC(n=2000,d=2)	1000	44	42	$0.102(\pm 0.007)$	$0.099(\pm 0.007)$	$0.187(\pm 0.050)$	$0.120(\pm 0.034)$
TS(n=2000,d=2)	1000	47	47	$0.101(\pm 0.007)$	$0.099(\pm 0.007)$	$0.147(\pm 0.032)$	$0.103(\pm 0.026)$
INC(n=2000,d=10)	1000	116	113	$0.511(\pm 0.016)$	$0.491(\pm 0.016)$	$0.611(\pm 0.052)$	$0.437(\pm 0.042)$
HCC(n=2000,d=10)	1000	110	105	$0.500(\pm 0.016)$	$0.481(\pm 0.015)$	$0.602(\pm 0.056)$	$0.430(\pm 0.044)$
TS(n=2000,d=10)	1000	101	98	$0.511(\pm 0.016)$	$0.499(\pm 0.016)$	$0.601(\pm 0.054)$	$0.486(\pm 0.051)$

Table G.3: Comparison of the min-max truncated estimator $\hat{g}$ with the ordinary least squares estimator $\hat{g}^{(\text{ols})}$ with the heavy-tailed noise (see Section G.1).

	nb of iterations	nb of iter. with $R(\hat{g}) \neq R(\hat{g}^{(\text{ols})})$	nb of iter. with $R(\hat{g}) < R(\hat{g}^{(\text{ols})})$	$\mathbb{E}R(\hat{g}^{(\text{ols})}) - R(g^*_{\mathbf{L}})$	$\mathbb{E}R(\hat{g}) - R(g^*_{\mathbf{L}})$	$\mathbb{E}R[\hat{g}^{(\text{ols})}]\|\hat{g} \neq \hat{g}^{(\text{ols})}] - R(g^*_{\mathbf{L}})$	$\mathbb{E}[R(\hat{g})\|\hat{g} \neq \hat{g}^{(\text{ols})}] - R(g^*_{\mathbf{L}})$
INC(n=200,d=1)	1000	163	145	$7.72(\pm 3.46)$	$3.92(\pm 0.409)$	$30.52(\pm 20.8)$	$7.20(\pm 1.61)$
INC(n=200,d=2)	1000	104	98	$22.69(\pm 23.14)$	$19.18(\pm 23.09)$	$45.36(\pm 14.1)$	$11.63(\pm 2.19)$
HCC(n=200,d=2)	1000	120	117	$18.16(\pm 12.68)$	$8.07(\pm 0.718)$	$99.39(\pm 105)$	$15.34(\pm 4.41)$
TS(n=200,d=2)	1000	110	105	$43.89(\pm 63.79)$	$39.71(\pm 63.76)$	$48.55(\pm 18.4)$	$10.59(\pm 2.01)$
INC(n=1000,d=2)	1000	104	100	$3.98(\pm 2.25)$	$1.78(\pm 0.128)$	$23.18(\pm 21.3)$	$2.03(\pm 0.56)$
INC(n=1000,d=10)	1000	253	242	$16.36(\pm 5.10)$	$7.90(\pm 0.278)$	$41.25(\pm 19.8)$	$7.81(\pm 0.69)$
HCC(n=1000,d=10)	1000	220	211	$13.57(\pm 1.93)$	$7.88(\pm 0.255)$	$33.13(\pm 8.2)$	$7.28(\pm 0.59)$
TS(n=1000,d=10)	1000	214	211	$18.67(\pm 11.62)$	$13.79(\pm 11.52)$	$30.34(\pm 7.2)$	$7.53(\pm 0.58)$
INC(n=2000,d=2)	1000	113	103	$1.56(\pm 0.41)$	$0.89(\pm 0.059)$	$6.74(\pm 3.4)$	$0.86(\pm 0.18)$
HCC(n=2000,d=2)	1000	105	97	$1.66(\pm 0.43)$	$0.95(\pm 0.062)$	$7.87(\pm 3.8)$	$1.13(\pm 0.23)$
TS(n=2000,d=2)	1000	101	95	$1.59(\pm 0.64)$	$0.88(\pm 0.058)$	$8.03(\pm 6.2)$	$1.04(\pm 0.22)$
INC(n=2000,d=10)	1000	259	255	$8.77(\pm 4.02)$	$4.23(\pm 0.154)$	$21.54(\pm 15.4)$	$4.03(\pm 0.39)$
HCC(n=2000,d=10)	1000	250	242	$6.98(\pm 1.17)$	$4.13(\pm 0.127)$	$15.35(\pm 4.5)$	$3.94(\pm 0.25)$
TS(n=2000,d=10)	1000	238	233	$8.49(\pm 3.61)$	$5.95(\pm 3.486)$	$14.82(\pm 3.8)$	$4.17(\pm 0.30)$

Table G.4: Comparison of the min-max truncated estimator $\hat{g}$ with the ordinary least squares estimator $\hat{g}^{(\text{ols})}$ with the asymmetric heavy-tailed noise (see Section G.1).

	nb of iterations	nb of iter. with $R(\hat{g}) \neq R(\hat{g}^{(\text{ols})})$	nb of iter. with $R(\hat{g}) < R(\hat{g}^{(\text{ols})})$	$\mathbb{E}R(\hat{g}^{(\text{ols})}) - R(g^*_{\mathbf{L}})$	$\mathbb{E}R(\hat{g}) - R(g^*_{\mathbf{L}})$	$\mathbb{E}R[(\hat{g}^{(\text{ols})})\|\hat{g} \neq \hat{g}^{(\text{ols})}] - R(g^*_{\mathbf{L}})$	$\mathbb{E}[R(\hat{g})\|\hat{g} \neq \hat{g}^{(\text{ols})}] - R(g^*_{\mathbf{L}})$
INC(n=200,d=1)	1000	87	77	$5.49(\pm 3.07)$	$3.00(\pm 0.330)$	$35.44(\pm 34.7)$	$6.85(\pm 2.48)$
INC(n=200,d=2)	1000	70	66	$19.25(\pm 23.23)$	$17.4(\pm 23.2)$	$37.95(\pm 13.1)$	$11.05(\pm 2.87)$
HCC(n=200,d=2)	1000	67	66	$7.19(\pm 0.88)$	$5.81(\pm 0.397)$	$31.52(\pm 10.5)$	$10.87(\pm 2.64)$
TS(n=200,d=2)	1000	76	68	$39.80(\pm 64.09)$	$37.9(\pm 64.1)$	$34.28(\pm 14.8)$	$9.21(\pm 2.05)$
INC(n=1000,d=2)	1000	101	92	$2.81(\pm 2.21)$	$1.31(\pm 0.106)$	$16.76(\pm 21.8)$	$1.88(\pm 0.69)$
INC(n=1000,d=10)	1000	211	195	$10.71(\pm 4.53)$	$5.86(\pm 0.222)$	$29.00(\pm 21.3)$	$6.03(\pm 0.71)$
HCC(n=1000,d=10)	1000	197	185	$8.67(\pm 1.16)$	$5.81(\pm 0.177)$	$20.31(\pm 5.59)$	$5.79(\pm 0.43)$
TS(n=1000,d=10)	1000	258	233	$13.62(\pm 11.27)$	$11.3(\pm 11.2)$	$14.68(\pm 2.45)$	$5.60(\pm 0.36)$
INC(n=2000,d=2)	1000	106	92	$1.04(\pm 0.37)$	$0.64(\pm 0.042)$	$4.54(\pm 3.45)$	$0.79(\pm 0.16)$
HCC(n=2000,d=2)	1000	99	90	$0.90(\pm 0.11)$	$0.66(\pm 0.042)$	$3.23(\pm 0.93)$	$0.82(\pm 0.16)$
TS(n=2000,d=2)	1000	84	81	$1.11(\pm 0.66)$	$0.60(\pm 0.042)$	$6.80(\pm 7.79)$	$0.69(\pm 0.17)$
INC(n=2000,d=10)	1000	238	222	$6.32(\pm 4.18)$	$3.07(\pm 0.147)$	$16.84(\pm 17.5)$	$3.18(\pm 0.51)$
HCC(n=2000,d=10)	1000	221	203	$4.49(\pm 0.98)$	$2.98(\pm 0.091)$	$9.76(\pm 4.39)$	$2.93(\pm 0.22)$
TS(n=2000,d=10)	1000	412	350	$5.93(\pm 3.51)$	$4.59(\pm 3.44)$	$6.07(\pm 1.76)$	$2.84(\pm 0.16)$

Table G.5: Comparison of the min-max truncated estimator $\hat{g}$ with the ordinary least squares estimator $\hat{g}^{(\text{ols})}$ for standard Gaussian noise.

	nb of iter.	nb of iter. with $R(\hat{g}) \neq R(\hat{g}^{(\text{ols})})$	nb of iter. with $R(\hat{g}) < R(\hat{g}^{(\text{ols})})$	$\mathbb{E}R(\hat{g}^{(\text{ols})}) - R(g^*_{\mathbf{L}})$	$\mathbb{E}R(\hat{g}) - R(g^*_{\mathbf{L}})$	$\mathbb{E}R[(\hat{g}^{(\text{ols})})\|\hat{g} \neq \hat{g}^{(\text{ols})}] - R(g^*_{\mathbf{L}})$	$\mathbb{E}[R(\hat{g})\|\hat{g} \neq \hat{g}^{(\text{ols})}] - R(g^*_{\mathbf{L}})$
INC(n=200,d=1)	1000	20	8	$0.541(\pm 0.048)$	$0.541(\pm 0.048)$	$0.401(\pm 0.168)$	$0.397(\pm 0.167)$
INC(n=200,d=2)	1000	1	0	$1.051(\pm 0.067)$	$1.051(\pm 0.067)$	2.566	2.757
HCC(n=200,d=2)	1000	1	0	$1.051(\pm 0.067)$	$1.051(\pm 0.067)$	2.566	2.757
TS(n=200,d=2)	1000	0	0	$1.068(\pm 0.067)$	$1.068(\pm 0.067)$	–	–
INC(n=1000,d=2)	1000	0	0	$0.203(\pm 0.013)$	$0.203(\pm 0.013)$	–	–
INC(n=1000,d=10)	1000	0	0	$1.023(\pm 0.029)$	$1.023(\pm 0.029)$	–	–
HCC(n=1000,d=10)	1000	0	0	$1.023(\pm 0.029)$	$1.023(\pm 0.029)$	–	–
TS(n=1000,d=10)	1000	0	0	$0.997(\pm 0.028)$	$0.997(\pm 0.028)$	–	–
INC(n=2000,d=2)	1000	0	0	$0.112(\pm 0.007)$	$0.112(\pm 0.007)$	–	–
HCC(n=2000,d=2)	1000	0	0	$0.112(\pm 0.007)$	$0.112(\pm 0.007)$	–	–
TS(n=2000,d=2)	1000	0	0	$0.098(\pm 0.006)$	$0.098(\pm 0.006)$	–	–
INC(n=2000,d=10)	1000	0	0	$0.517(\pm 0.015)$	$0.517(\pm 0.015)$	–	–
HCC(n=2000,d=10)	1000	0	0	$0.517(\pm 0.015)$	$0.517(\pm 0.015)$	–	–
TS(n=2000,d=10)	1000	0	0	$0.501(\pm 0.015)$	$0.501(\pm 0.015)$	–	–

Figure G.1: Surrounding points are the points of the training set generated several times from $TS(1000, 10)$ (with the mixture noise with $p = 0.005$ and $\rho = 0.4$) that are not taken into account in the min-max truncated estimator (to the extent that the estimator would not change by removing simultaneously all these points). The min-max truncated estimator $x \mapsto \hat{f}(x)$ appears in dash-dot line, while $x \mapsto \mathbb{E}(Y|X = x)$ is in solid line. In these six simulations, it outperforms the ordinary least squares estimator.

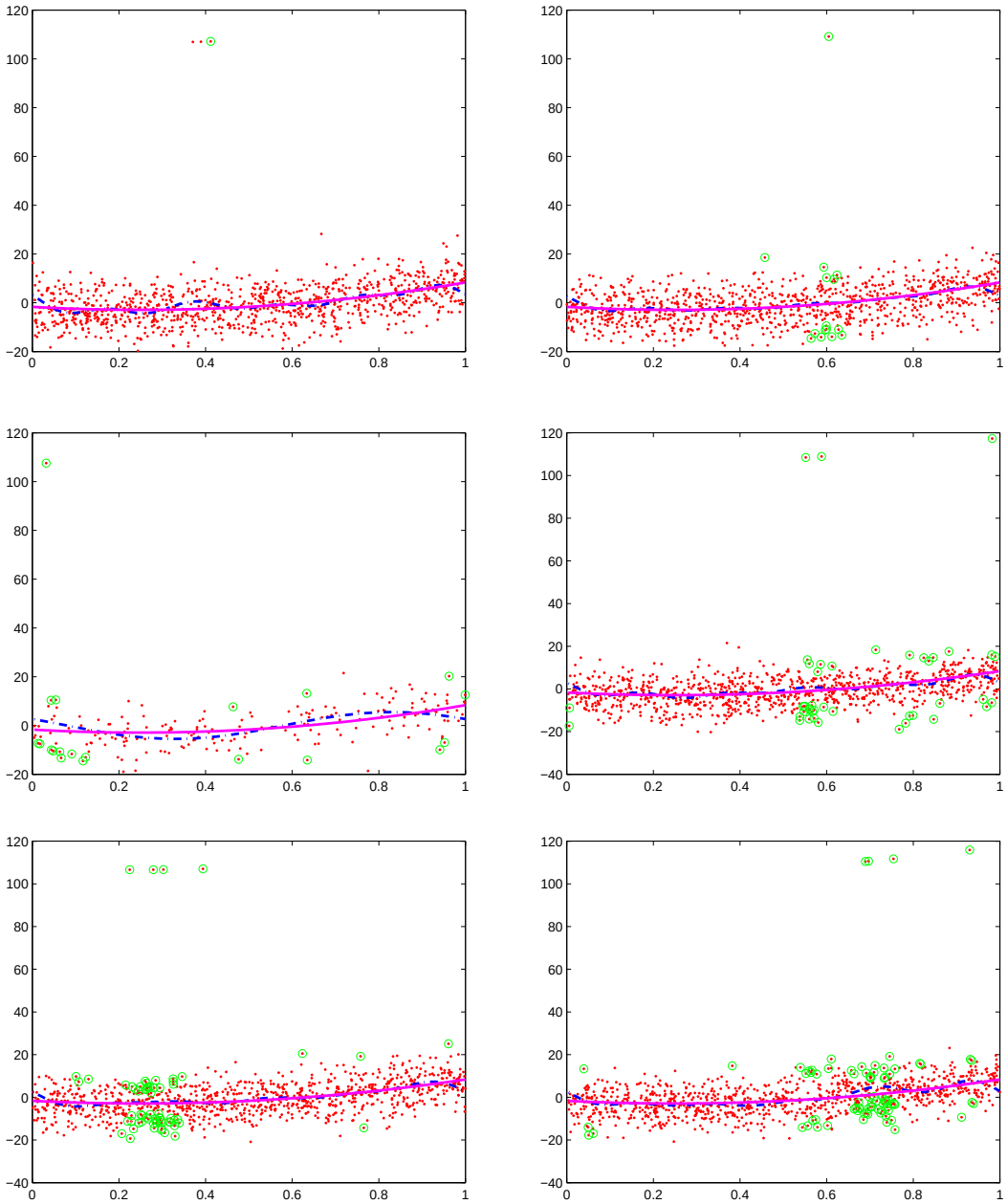
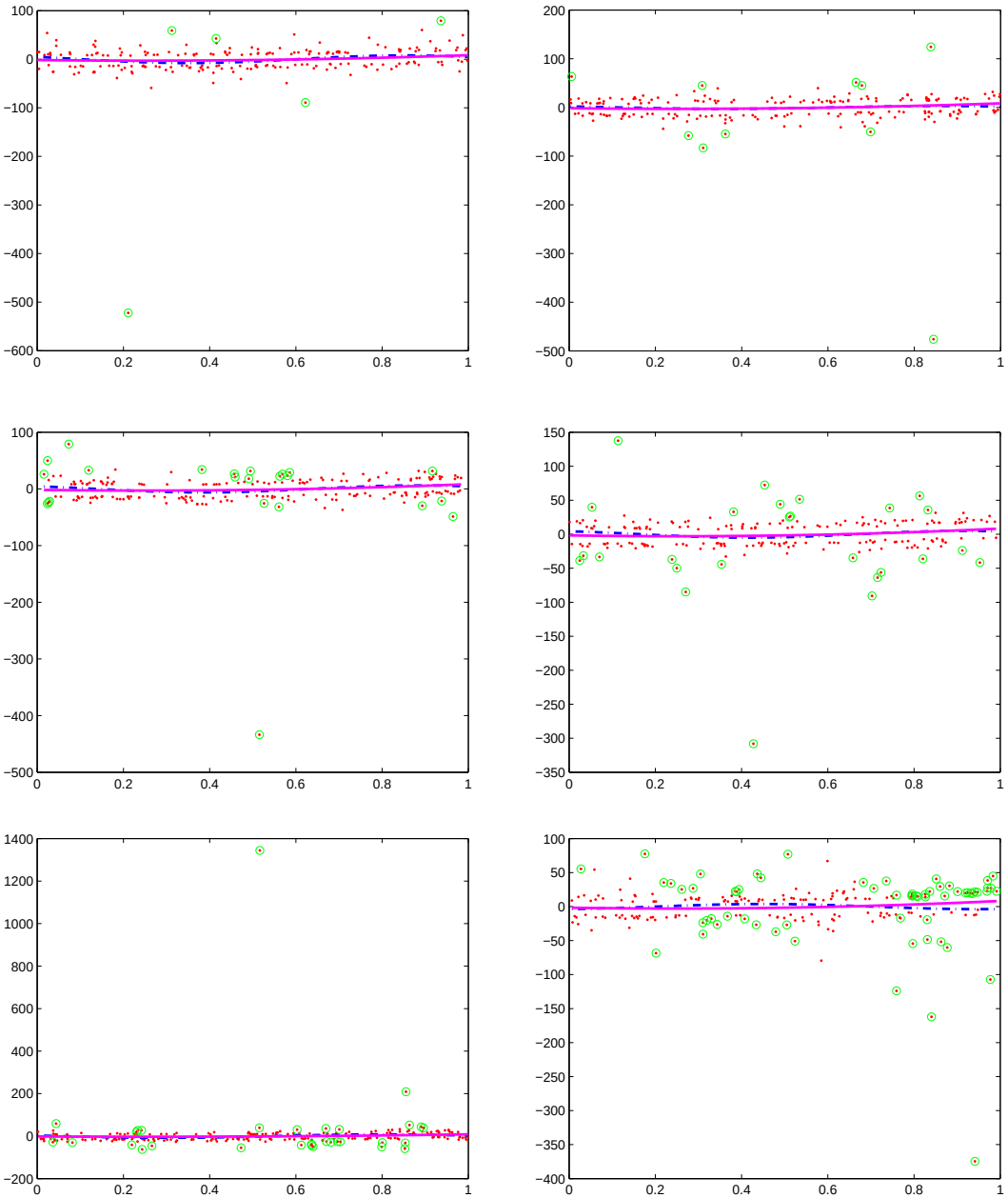


Figure G.2: Surrounding points are the points of the training set generated several times from $TS(200, 2)$ (with the heavy-tailed noise) that are not taken into account in the min-max truncated estimator (to the extent that the estimator would not change by removing these points). The min-max truncated estimator $x \mapsto \hat{f}(x)$ appears in dash-dot line, while $x \mapsto \mathbb{E}(Y|X = x)$ is in solid line. In these six simulations, it outperforms the ordinary least squares estimator. Note that in the last figure, it does not consider 64 points among the 200 training points.



Bibliography

[1] R. Agrawal. The continuum-armed bandit problem. *SIAM J. Control and Optimization*, 33:1926–1951, 1995.

[2] R. Agrawal. Sample mean based index policies with o(log n) regret for the multi-armed bandit problem. *Advances in Applied Mathematics*, 27:1054–1078, 1995.

[3] H. Akaike. Information theory and an extension of the maximum likelihood principle. *Budapest, Hungary*, pages 267–281, 1973.

[4] P. Alquier. *Transductive and inductive adaptative inference for regression and density estimation*. PhD thesis, PhD thesis, University Paris 6, 2006.

[5] P. Alquier. PAC-Bayesian bounds for randomized empirical risk minimizers. *Mathematical Methods of Statistics*, 17(4):279–304, 2008.

[6] A. Ambroladze, E. Parrado-Hernández, and J. Shawe-Taylor. Tighter PAC-bayes bounds. In *Advances in Neural Information Processing Systems 18*, pages 9–16, 2006.

[7] J.-Y. Audibert. *PAC-Bayesian statistical learning theory*. PhD thesis, Laboratoire de Probabilités et Modèles Aléatoires, Universités Paris 6 and Paris 7, 2004. `http://imagine.enpc.fr/~audibert/ThesePack.zip`.

[8] J.-Y. Audibert. Progressive mixture rules are deviation suboptimal. *Advances in Neural Information Processing Systems*, 20, 2007.

[9] J.-Y. Audibert. Fast learning rates in statistical inference through aggregation. *Ann. Stat.*, 37:1591–1646, 2009.

[10] J.-Y. Audibert and O. Bousquet. Combining PAC-bayesian and generic chaining bounds. *Journal of Machine Learning Research*, 8:863–889, 2007.

[11] J.-Y. Audibert and S. Bubeck. Minimax policies for adversarial and stochastic bandits. 2009.

[12] J.-Y. Audibert and S. Bubeck. Regret bounds and minimax policies under partial monitoring. *Journal of Machine Learning Research*, 2010.

[13] J.-Y. Audibert, S. Bubeck, and R. Munos. Best Arm Identification in Multi-Armed Bandits. In *Proceedings of the 23th annual conference on Computational Learning Theory (COLT)*, 2010.

[14] J.-Y. Audibert and O. Catoni. Risk bounds in linear regression through PAC-bayesian truncation. Technical report, Feb 2009.

[15] J.-Y. Audibert, R. Munos, and Cs. Szepesvri. Exploration-exploitation trade-off using variance estimates in multi-armed bandits. *Theoretical Computer Science*, 410:1876–1902, 2009.

[16] J.-Y. Audibert and A.B. Tsybakov. Fast learning rates for plug-in classifiers. *Ann. Statist.*, 35(2):608–633, 2007.

[17] P. Auer. Using confidence bounds for exploitation-exploration trade-offs. *Journal of Machine Learning Research*, 3:397–422, 2002.

[18] P. Auer, N. Cesa-Bianchi, and P. Fischer. Finite-time analysis of the multi-armed bandit problem. *Mach. Learn.*, 47(2-3):235–256, 2002.

[19] P. Auer, N. Cesa-Bianchi, Y. Freund, and R. Schapire. The nonstochastic multi-armed bandit problem. *SIAM Journal on Computing*, 32(1):48–77, 2002.

[20] P. Auer, R. Ortner, and Cs. Szepesvári. Improved rates for the stochastic continuum-armed bandit problem. *20th COLT, San Diego, CA, USA*, 2007.

[21] M. Babaioff, Y. Sharma, and A. Slivkins. Characterizing truthful multi-armed bandit mechanisms: extended abstract. In *Proceedings of the tenth ACM conference on Electronic commerce*, pages 79–88. ACM, 2009.

[22] F.R. Bach. Consistency of the group Lasso and multiple kernel learning. *Journal of Machine Learning Research*, 9:1179–1225, 2008.

[23] F.R. Bach, G.R.G. Lanckriet, and M.I. Jordan. Multiple kernel learning, conic duality, and the SMO algorithm. In *Proceedings of the twenty-first international conference on Machine learning*, 2004.

[24] A. Barron. Are bayes rules consistent in information? In T.M. Cover and B. Gopinath, editors, *Open Problems in Communication and Computation*, pages 85–91. Springer, 1987.

[25] A. Barron and Y. Yang. Information-theoretic determination of minimax rates of convergence. *Ann. Stat.*, 27(5):1564–1599, 1999.

[26] P.L. Bartlett, O. Bousquet, and S. Mendelson. Local rademacher complexities. *Annals of Statistics*, 33(4):1497–1537, 2005.

[27] P.L. Bartlett, M.I. Jordan, and J.D. McAuliffe. Convexity, classification, and risk bounds. *Journal of the American Statistical Association*, 101:138–156, 2006.

[28] P.L. Bartlett and M. Traskin. Adaboost is consistent. *J. Mach. Learn. Res.*, 8:2347–2368, 2007.

[29] D. Bergemann and J. Valimaki. Bandit problems. 2008. In The New Palgrave Dictionary of Economics, 2nd ed. Macmillan Press.

[30] D. A. Berry, R. W. Chen, A. Zame, D. C. Heath, and L. A. Shepp. Bandit problems with infinitely many arms. *Annals of Statistics*, 25(5):2103–2116, 1997.

[31] L. Birgé and P. Massart. Minimum contrast estimators on sieves: exponential bounds and rates of convergence. *Bernoulli*, 4(3):329–375, 1998.

[32] L. Birgé and P. Massart. Minimal penalties for Gaussian model selection. *Probability Theory and Related Fields*, 138(1):33–73, 2007.

[33] G. Blanchard. *Méthodes de mélange et d'agrégation d'estimateurs en reconnaissance de formes. Application aux arbres de décision*. PhD thesis, Université Paris 13 – Paris Nord, 2001.

[34] S. Boucheron, O. Bousquet, and G. Lugosi. Theory of classification: some recent advances. *ESAIM Probability & Statistics*, 9:323–375, 2005.

[35] S. Bubeck, R. Munos, G. Stoltz, and C. Szepesvari. Online optimization in X-armed bandits. In *Advances in Neural Information Processing Systems 21*, pages 201–208. 2009.

[36] F. Bunea, A.B. Tsybakov, and M.H. Wegkamp. Aggregation for Gaussian regression. *Annals of Statistics*, 35(4), 2007.

[37] O. Catoni. *Statistical learning theory and stochastic optimization*. Springer. Probability summer school, Saint Flour 2001.

[38] O. Catoni. A mixture approach to universal model selection. preprint LMENS 97-30, Available from `http://www.dma.ens.fr/edition/preprints/Index.97.html`, 1997.

[39] O. Catoni. Universal aggregation rules with exact bias bound. Preprint n.510, `http://www.proba.jussieu.fr/mathdoc/preprints/`, 1999.

[40] O. Catoni. A PAC-Bayesian approach to adaptive classification. Preprint n.840, Laboratoire de Probabilités et Modèles Aléatoires, Universités Paris 6 and Paris 7, 2003.

[41] O. Catoni. *PAC-Bayesian supervised classification: the thermodynamics of statistical learning*. Lecture Notes series of the IMS, 2007.

[42] O. Catoni. High confidence estimates of the mean of heavy-tailed real random variables. 2009. Available on Arxiv.

[43] N. Cesa-Bianchi. Analysis of two gradient-based algorithms for on-line regression. *J. Comput. Syst. Sci*, 59(3):392–411, 1999.

[44] N. Cesa-Bianchi, Y. Freund, D. Haussler, D.P. Helmbold, R.E. Schapire, and M.K. Warmuth. How to use expert advice. *J. ACM*, 44(3):427–485, 1997.

[45] N. Cesa-Bianchi and G. Lugosi. On prediction of individual sequences. *Ann. Stat.*, 27(6):1865–1895, 1999.

[46] N. Cesa-Bianchi and G. Lugosi. *Prediction, Learning, and Games*. Cambridge University Press, 2006.

[47] N. Cesa-Bianchi, G. Lugosi, and G. Stoltz. Minimizing regret with label efficient prediction. *IEEE Transactions on Information Theory*, 51(6):2152–2162, 2005.

[48] P.A. Coquelin and R. Munos. Bandit algorithms for tree search. In *Uncertainty in Artificial Intelligence*, 2007.

[49] Paul Dagum, Richard Karp, Michael Luby, and Sheldon Ross. An optimal algorithm for Monte Carlo estimation. *SIAM Journal on Computing*, 29(5):1484–1496, 2000.

[50] A. Dalalyan and A. Tsybakov. Aggregation by exponential weighting, sharp oracle inequalities and sparsity. *Machine Learning*, 72:39–61, 2008.

[51] A. Dalalyan and A. Tsybakov. Sparse regression learning by aggregation and langevin monte-carlo. In *22nd Annual Conference on Learning Theory*, Montreal, Canada, Jun 2009.

[52] N.R. Devanur and S.M. Kakade. The price of truthfulness for pay-per-click auctions. In *Proceedings of the tenth ACM conference on Electronic commerce*, pages 99–106. ACM, 2009.

[53] L. Devroye, L. Györfi, and G. Lugosi. *A Probabilistic Theory of Pattern Recognition*. Springer-Verlag, 1996.

[54] C. Domingo, R. Gavaldà, and O. Watanabe. Adaptive sampling methods for scaling up knowledge discovery algorithms. *Data Mining and Knowledge Discovery*, 6(2):131–152, 2002.

[55] R.M. Dudley. Central limit theorems for empirical measures. *Ann. Probab.*, 6:899–929, 1978.

[56] E. Even-Dar, S. Mannor, and Y. Mansour. Action elimination and stopping conditions for the multi-armed bandit and reinforcement learning problems. *The Journal of Machine Learning Research*, 7:1079–1105, 2006.

[57] D.A. Freedman. On tail probabilities for martingales. *The Annals of Probability*, 3(1):100–118, 1975.

[58] Y. Freund and R. E. Schapire. A decision–theoretic generalization of on–line learning and an application to boosting. *Journal of Computer and System Sciences*, 55(1):119–139, August 1997.

[59] S. Gelly and Y. Wang. Exploration exploitation in go: UCT for Monte-Carlo go. In *Online trading between exploration and exploitation Workshop, Twentieth Annual Conference on Neural Information Processing Systems (NIPS 2006)*, 2006.

[60] P. Germain, A. Lacasse, F. Laviolette, and M. Marchand. PAC-Bayesian learning of linear classifiers. In *Proceedings of the 26th Annual International Conference on Machine Learning (ICML)*, pages 353–360, 2009.

[61] J. C. Gittins. *Multi-armed Bandit Allocation Indices*. Wiley-Interscience series in systems and optimization. Wiley, Chichester, NY, 1989.

[62] S. Grünewälder, J.-Y. Audibert, M. Opper, and J. Shawe-Taylor. Regret bounds for gaussian process bandit problems. In *Proceedings of the 14th International Conference on Artificial Intelligence and Statistics (AISTATS)*, 2010.

[63] L. Györfi, M. Kohler, A. Krzyżak, and H. Walk. *A Distribution-Free Theory of Nonparametric Regression*. Springer, 2004.

[64] A. György and G. Ottucsák. Adaptive routing using expert advice. *Computer Journal-Oxford*, 49(2):180–189, 2006.

[65] D. Haussler, J. Kivinen, and M. K. Warmuth. Sequential prediction of individual sequences under general loss functions. *IEEE Trans. on Information Theory*, 44(5):1906–1925, 1998.

[66] D. Helmbold and S. Panizza. Some label efficient learning results. In *Proceedings of the 10th annual conference on Computational learning theory*, pages 218–230. ACM New York, NY, USA, 1997.

[67] W. Hoeffding. Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association*, 58:13–30, 1963.

[68] J.H. Holland. *Adaptation in natural and artificial systems*. MIT press Cambridge, MA, 1992.

[69] J. Huang and T. Zhang. The benefit of group sparsity. 2009. Available on Arxiv.

[70] R. Jenatton, J.-Y. Audibert, and F. Bach. Structured variable selection with sparsity-inducing norms. 2009. Available on Arxiv.

[71] A. Juditsky and A. Nemirovski. Functional aggregation for nonparametric estimation. *Ann. Stat.*, 28:681–712, 2000.

[72] A. Juditsky, P. Rigollet, and A.B. Tsybakov. Learning by mirror averaging. *Ann. Statist*, 36(5):2183–2206, 2008.

[73] A.B. Juditsky, A.V. Nazin, A.B. Tsybakov, and N. Vayatis. Recursive aggregation of estimators by the mirror descent algorithm with averaging. *Problems of Information Transmission*, 41(4):368–384, 2005.

[74] J. Kivinen and M.K. Warmuth. Exponentiated gradient versus gradient descent for linear predictors. *Information and Computation*, 1997.

[75] R. Kleinberg, A. Slivkins, and E. Upfal. Multi-armed bandit problems in metric spaces. In *Proceedings of the 40th ACM Symposium on Theory of Computing*, 2008.

[76] R. Kleinberg, A. Slivkins, and E. Upfal. Multi-armed bandits in metric spaces. In *Proceedings of the 40th annual ACM symposium on Theory of computing*, pages 681–690, 2008.

[77] R. D. Kleinberg. Nearly tight bounds for the continuum-armed bandit problem. In *Advances in Neural Information Processing Systems 17*, pages 697–704. 2005.

[78] R.D. Kleinberg. Nearly tight bounds for the continuum-armed bandit problem. In *NIPS-2004*, 2004.

[79] L. Kocsis and Cs. Szepesvári. Bandit based Monte-Carlo planning. In *Proceedings of the 17th European Conference on Machine Learning (ECML-2006)*, pages 282–293, 2006.

[80] V. Koltchinskii. Local rademacher complexities and oracle inequalities in risk minimization. *Annals of Statistics*, 34(6), 2006.

[81] V. Koltchinskii and M. Yuan. Sparse recovery in large ensembles of kernel machines. In *Conference on Learning Theory, COLT*, pages 229–238, 2008.

[82] A. Lacasse, F. Laviolette, and M. Marchand. PAC-Bayesian Learning of Linear Classifiers. In *Proceedings of the 26th International Conference on Machine Learning*, 2009.

[83] A. Lacasse, F. Laviolette, M. Marchand, P. Germain, and N. Usunier. PAC-Bayes Bounds for the Risk of the Majority Vote and the Variance of the Gibbs Classifier. *Advances in Neural Information Processing Systems*, 19:769, 2007.

[84] T. L. Lai and H. Robbins. Asymptotically efficient adaptive allocation rules. *Advances in Applied Mathematics*, 6:4–22, 1985.

[85] D. Lamberton, G. Pagès, and P. Tarrès. When can the two-armed bandit algorithm be trusted? *Annals of Applied Probability*, 14(3):1424–1454, 2004.

[86] G.R.G. Lanckriet, N. Cristianini, P. Bartlett, L.E. Ghaoui, and M.I. Jordan. Learning the kernel matrix with semidefinite programming. *Journal of Machine Learning Research*, 5:27–72, 2004.

[87] J. Langford, M. Seeger, and N. Megiddo. An improved predictive accuracy bound for averaging classifiers. In *Proceedings of the 18th International Conference on Machine Learning*, pages 290–297, 2001.

[88] J. Langford and J. Shawe-Taylor. PAC-Bayes & margins. *Advances in neural information processing systems*, pages 439–446, 2003.

[89] F. Laviolette and M. Marchand. PAC-Bayes risk bounds for stochastic averages and majority votes of sample-compressed classifiers. *Journal of Machine Learning Research*, 8:1461–1487, 2007.

[90] G. Lecué. Suboptimality of penalized empirical risk minimization in classification. In *Proceedings of the 20th annual conference on Computational Learning Theory (COLT), Lecture Notes in Computer Science*, volume 4539, pages 142 – 156, 2007.

[91] G. Lecué and S. Mendelson. Aggregation via empirical risk minimization. *Probability Theory and Related Fields*, 145(3):591–613, 2009.

[92] W.S. Lee, P.L. Bartlett, and R.C. Williamson. The importance of convexity in learning with squared loss. *IEEE Trans. Inform. Theory*, 44(5):1974–1980, 1998.

[93] K. Lounici. Sup-norm convergence rate and sign concentration property of Lasso and Dantzig estimators. *Electronic Journal of Statistics*, 2:90–102, 2008.

[94] K. Lounici, M. Pontil, A.B. Tsybakov, and S. van de Geer. Taking advantage of sparsity in multi-task learning. In *Proceedings of the 22th annual conference on Computational Learning Theory (COLT)*, 2009.

[95] G. Lugosi and N. Vayatis. On the bayes-risk consistency of regularized boosting methods. *Ann. Stat.*, 32(1):30–55, 2004.

[96] C. L. Mallows. Some comments on Cp. *Technometrics*, 15:661–675, 1973.

[97] E. Mammen and A.B. Tsybakov. Smooth discrimination analysis. *Ann. Stat.*, 27:1808–1829, 1999.

[98] O. Maron and A. W. Moore. Hoeffding races: Accelerating model selection search for classification and function approximation. In *NIPS*, pages 59–66, 1993.

[99] P. Massart. Some applications of concentration inequalities to statistics. *Ann. Fac. Sci. Toulouse*, Math. 9(2):245–303, 2000.

[100] P. Massart and E. Nédélec. Risk bounds for statistical learning. *Ann. Statist.*, 34(5):2326–2366, 2006.

[101] A. Maurer and M. Pontil. Empirical Bernstein Bounds and Sample Variance Penalization. *stat*, 1050:21, 2009.

[102] D. A. McAllester. PAC-Bayesian model averaging. In *Proceedings of the 12th annual conference on Computational Learning Theory*, pages 164–170, 1999.

[103] D. A. McAllester. Simplified PAC-bayesian margin bounds. In *COLT: Proceedings of the Workshop on Computational Learning Theory*, 2003.

[104] L. Meier, S. Van de Geer, and P. Bühlmann. High-dimensional additive modeling. *Annals of Statistics*, 37:3779–3821, 2009.

[105] N. Meinshausen and B. Yu. Lasso-type recovery of sparse representations for high-dimensional data. *Annals of Statistics*, 37(1):246–270, 2009.

[106] S. Mendelson. Lower bounds for the empirical minimization algorithm. *Information Theory, IEEE Transactions on*, 54(8):3797–3803, 2008.

[107] N. Merhav and M. Feder. Universal prediction. *IEEE Transactions on Information Theory*, 44(6):2124–2147, 1998.

[108] C.A. Micchelli and M. Pontil. Learning the kernel function via regularization. *Journal of Machine Learning Research*, 6(2):1099, 2006.

[109] V. Mnih, Cs. Szepesvári, and J.-Y. Audibert. Empirical bernstein stopping. In *Proceedings of the 25th International Conference (ICML)*, volume 307, pages 672–679, 2008.

[110] A. Nemirovski. *Lectures on probability theory and statistics. Part II: topics in Non-parametric statistics*. Springer-Verlag. Probability summer school, Saint Flour 1998.

[111] C.S. Ong, A.J. Smola, and R.C. Williamson. Learning the kernel with hyperkernels. *Journal of Machine Learning Research*, 6:1043–1071, 2005.

[112] L.E. Ortiz and L.P. Kaelbling. Sampling methods for action selection in influence diagrams. In *AAAI/IAAI*, pages 378–385, 2000.

[113] M.R. Osborne, B. Presnell, and B.A. Turlach. On the lasso and its dual. *Journal of Computational and Graphical Statistics*, pages 319–337, 2000.

[114] G. Raskutti, M.J. Wainwright, and B. Yu. Minimax rates of estimation for high-dimensional linear regression over lq-balls. *ArXiv*, 910, 2009.

[115] H. Robbins. Some aspects of the sequential design of experiments. *Bulletin of the American Mathematics Society*, 58:527–535, 1952.

[116] G. Schwarz. Estimating the dimension of a model. *The annals of statistics*, pages 461–464, 1978.

[117] M. Seeger. PAC-Bayesian generalization error bounds for gaussian process classification. Informatics report series EDI-INF-RR-0094, Division of Informatics, University of Edinburgh, 2002.

[118] Y. Seldin and N. Tishby. Multi-classification by categorical features via clustering. In *Proceedings of the 25th international conference on Machine learning*, pages 920–927, 2008.

[119] R. S. Sutton and A. G. Barto. *Reinforcement Learning: An Introduction*. MIT Press, 1998.

[120] M. Talagrand. Majorizing chaining: the generic chaining. *Ann. Probab.*, 24:1049–1103, 1996.

[121] O. Teytaud, S. Gelly, and M. Sebag. Anytime many-armed bandit. *Conférence francophone sur l'Apprentissage automatique (CAp) Grenoble, France*, 2007.

[122] R. Tibshirani. Regression shrinkage and selection via the lasso. *J. Roy. Stat. Soc. B*, 58:267–288, 1994.

[123] A.B. Tsybakov. Optimal rates of aggregation. In *Computational Learning Theory and Kernel Machines, Lecture Notes in Artificial Intelligence*, volume 2777, pages 303–313, 2003.

[124] S.A. Van De Geer. High-dimensional generalized linear models and the lasso. *Annals of Statistics*, 36(2):614, 2008.

[125] V. Vapnik. *Estimation of Dependences Based on Empirical Data*. Springer-Verlag, Berlin, 1982.

[126] V.G. Vovk. Aggregating strategies. In *COLT '90: Proceedings of the third annual workshop on Computational learning theory*, pages 371–386, 1990.

[127] V.G. Vovk. A game of prediction with expert advice. *Journal of Computer and System Sciences*, pages 153–173, 1998.

[128] M. J. Wainwright. Sharp thresholds for noisy and high-dimensional recovery of sparsity. *IEEE transactions on information theory*, 55(5):2183–2202, 2009.

[129] Y. Wang, J.-Y. Audibert, and R. Munos. Algorithms for infinitely many-armed bandits. *Advances in Neural Information Processing Systems (NIPS)*, 21:1729–1736, 2008.

[130] O. Watanabe. Simple sampling techniques for discovery science. *IEICE Transactions on Information and Systems*, 1:19–26, 2000.

[131] Y. Yang. Combining different procedures for adaptive regression. *Journal of multivariate analysis*, 74:135–161, 2000.

[132] Y. Yang. Adaptive regression by mixing. *Journal of American Statistical Association*, 96:574–588, 2001.

[133] Y. Yang. Aggregating regression procedures to improve performance. *Bernoulli*, 10:25–47, 2004.

[134] R. Yaroshinsky, R. El-Yaniv, and S.S. Seiden. How to better use expert advice. *Mach. Learn.*, 55(3):271–309, 2004.

[135] M. Yuan and Y. Lin. Model selection and estimation in regression with grouped variables. *Journal of the Royal Statistical Society Series B Statistical Methodology*, 68(1):49, 2006.

[136] T. Zhang. Information theoretical upper and lower bounds for statistical estimation. *IEEE Transaction on Information Theory*, 52(4):1307–1321, 2006.

[137] T. Zhang. From ϵ-entropy to KL-entropy: Analysis of minimum information complexity density estimation. *Annals of Statistics*, 34(5), 2007.

[138] P. Zhao and B. Yu. On model selection consistency of Lasso. *Journal of Machine Learning Research*, 7:2541–2563, 2006.