Thesis presented to Université Paris-Sud for
the degree of Habilitation

In Mathematics

Contributions to the sequential prediction of arbitrary sequences:

applications to the theory of repeated games and empirical studies of the performance of the aggregation of experts

by

Gilles Stoltz

CNRS researcher at Ecole normale supérieure, Paris
Affiliated professor at HEC Paris

Defended on February 3, 2011 before a jury composed of

Elisabeth	Gassiat	Université Paris-Sud	President
Gábor	Lugosi	ICREA / Universitat Pompeu Fabra	Examiner
Pascal	Massart	Université Paris-Sud	Reviewer
Eric	Moulines	Télécom ParisTech	Reviewer
Sylvain	Sorin	Université Pierre et Marie Curie	Examiner
Bruno	Sportisse	INRIA	Examiner

and after a public reading of the report also written by

Avrim	Blum	Carnegie Mellon University	Reviewer

Foreword and acknowledgements

Foreword

This manuscript was thought out and written in French. It summarizes the research I have been carrying since my early steps in 2002 —when I became a PhD student— to the present time. It is targeted to a wide audience of French mathematicians and computer scientists and is meant to be the main document in the file I had to complete to be awarded the habilitation (to supervise research works).

The present English version was translated from the original French version in the mere intention of sending out the manuscript to international reviewers —and I only had a short period of time to do so. In any case, my efforts were not vain if these lines are being read. Please accept my apologies for the remaining grammatical or stylistic mistakes!

Acknowledgements

Detailed acknowledgements can be found in the French version of the manuscript. I only reproduce below (parts of) the paragraphs intended therein to non-French-speaking colleagues.

Gábor, I already listed in the foreword to my thesis how nice and efficient a supervisor you had been. Thanks for continuing —after my thesis— to follow my advances and providing me with some useful recommendations, and even, some good research problems. One of my favorite post-thesis results was our elementary and constructive proof of the no-regret theorem of Rustichini's, which we exhibited jointly with Shie...

Speaking of Shie: I always appreciate your optimism. It is an excellent counterpart to my skepticism... You are such a terrific host, it is always a pleasure to visit you in Haifa!

Avrim, thanks for accepting to read and review this manuscript. A long and tough job for which you had to sacrifice a fraction of your Christmas vacations... You could not make it to the defense but my invitation to Paris still stands for future months (or years).

Nicolò, you too have been supportive at some key moments. It has been some time that we did not formally write a paper together, but we have been exchanging ideas and sharing our problems and solutions. It is always a pleasure to meet you at conferences, or in Paris or Milano.

Tomasz, you patiently taught me macro-economics and introduced to this whole new world called economics. Definitely, a different community! We worked hard, as the number of performed χ^2 tests shows, but also laughed several times thanks to your Polish sense of humor!

Publications

By type

The articles [1, 2, 3, 4], as well as a significant fraction of [5], were extracted from my PhD thesis. Other articles are posterior.

Articles in journals

[1] Gilles Stoltz and Gábor Lugosi. Internal regret in on-line portfolio selection. *Machine Learning*, 59:125–159, 2005.

[2] Nicolò Cesa-Bianchi, Gábor Lugosi, and Gilles Stoltz. Minimizing regret with label-efficient prediction. *IEEE: Transactions on Information Theory*, 51:2152–2162, 2005.

[3] Nicolò Cesa-Bianchi, Gábor Lugosi, and Gilles Stoltz. Regret minimization under partial monitoring. *Mathematics of Operations Research*, 31:562–580, 2006.

[4] Gilles Stoltz and Gábor Lugosi. Learning correlated equilibria in games with compact sets of strategies. *Games and Economic Behavior*, 59:187–208, 2007.

[5] Nicolò Cesa-Bianchi, Yishay Mansour, and Gilles Stoltz. Improved second-order inequalities for prediction under expert advice. *Machine Learning*, 66:321–352, 2007.

[6] Gábor Lugosi, Shie Mannor, and Gilles Stoltz. Strategies for prediction under imperfect monitoring. *Mathematics of Operations Research*, 33:513–528, 2008.

[7] Boris Mauricette, Vivien Mallet, and Gilles Stoltz. Ozone ensemble forecast with machine learning algorithms. *Journal of Geophysical Research*, 114:D05307, 2009.

[8] Sébastien Bubeck, Rémi Munos, and Gilles Stoltz. Pure exploration in finitely-armed and continuous-armed bandit problems. *Theoretical Computer Science*, 2010. In press.

[9] Shie Mannor and Gilles Stoltz. A geometric proof of calibration. *Mathematics of Operations Research*, 2010. In press.

[10] Gilles Stoltz. Agrégation séquentielle de prédicteurs : méthodologie générale et applications à la prévision de la qualité de l'air et à celle de la consommation électrique. *Journal de la Société Française de Statistique*, 151(2):66–106, 2010. (Survey paper following the award of the Marie-Jeanne Laurent-Duhamel prize).

Conference articles

[11] Sébastien Bubeck, Rémi Munos, Gilles Stoltz, and Csaba Szepesvári. Hierarchical optimistic optimization. In *Proceedings of NIPS'08*, pages 201–208. Curran Associates, Inc., 2008. (Extended version submitted for journal publication entitled: "$\mathcal{X}$–armed bandits").

[12] Gábor Lugosi, Omiros Papaspiliopoulos, and Gilles Stoltz. Online multi-task learning with hard constraints. In *Proceedings of COLT'09*. Omnipress, 2009.

Extended abstracts of the articles [1, 2, 5, 6, 8] were respectively published in the following conferences: COLT'03 ("Student paper award"), COLT'04, COLT'05 ("Student paper award"), COLT'07, and ALT'09.

Articles submitted for publication in journals

[13] Marie Devaine, Yannig Goude, and Gilles Stoltz. Forecasting the electricity consumption by aggregation of specialized experts; application to Slovakian and French country-wide hourly predictions. 2010.

[14] Tomasz Michalski and Gilles Stoltz. Do countries falsify economic data strategically? Some evidence that they do. 2010.

An extended version of the conference paper [11] is also currently considered by a journal.

Technical report and PhD thesis

The articles [7, 13, 10] rely on the technical reports [MMS07, GMS08, DGS09]; the latter are more detailed than the former in the sense that they report the empirical results of both the good and bad ideas considered. Their references are provided in the main bibliography, in which my thesis manuscript [Sto05] is also listed.

Textbook (in French)

[15] Vincent Rivoirard and Gilles Stoltz. *Statistique en action*. Vuibert, 2009. (Main volume is 320-page long and a supplementary 210-page long volume can be freely downloaded.).

By themes

1. Studies at the foundations of the prediction of individual sequences: [2, 5, 12]
2. Interactions with the theory of repeated games: [1, 3, 4, 6, 9]
3. Applications of the sequential aggregation aggregation of experts (methodological advances and empirical studies): [1, 7, 13, 10]
4. Stochastic continuum-armed bandits and miscellaneous works: [8, 11] and [14, 15]

By co-authors

Hosted by non-French academic institutions

Gábor Lugosi (ICREA and Universitat Pompeu Fabra, Barcelona): [1, 2, 3, 4, 6, 12]

Nicolò Cesa-Bianchi (Universitá degli studi, Milano): [2, 3, 5]

Yishay Mansour (Tel Aviv University): [5]

Shie Mannor (Mc Gill University, Montréal, and Israeli Institute of Technology, Technion): [6, 9]

Csaba Szepesvári (University of Alberta): [11]

Omiros Papaspiliopoulos (Universitat Pompeu Fabra, Barcelona): [12]

Hosted by French academic or R&D institutions

Vivien Mallet (INRIA, Paris-Rocquencourt) and Boris Mauricette (MSc student at University Paris-Diderot): [7]

Sébastien Bubeck and Rémi Munos (INRIA, Lille): [8, 11]

Yannig Goude (EDF R&D, Clamart) and Marie Devaine (MSc student at University Paris-Sud, Orsay): [13]

Vincent Rivoirard (University Paris-Sud, Orsay, and Ecole normale supérieure, Paris): [15]

Tomasz Michalski (HEC Paris): [14]

Curriculum vitæ

Academic positions held

Sept. 2007 – onwards	Affiliated professor at HEC[a] Paris
Oct. 2005 – onwards	CNRS researcher at Ecole normale supérieure, Paris
Sept. 2004 – Sept. 2005	Lecturer at Ecole normale supérieure, Paris
Sept. 2003 – Aug. 2004	Teaching assistant at Université Paris-Sud, Orsay

[a] A highly-ranked French business school

Curriculum

Sept. 2002 – May 2005	PhD student at Université Paris-Sud, Orsay; co-supervised by Gàbor Lugosi (Universitat Pompeu Fabra, Barcelona; main advisor) and Pascal Massart (Université Paris-Sud, Orsay; secondary advisor)
Sept. 1999 – Aug. 2003	BSc and MSc student (as a civil servant) at Ecole normale supérieure, Cachan

Awards and distinctions

June 2010	Best reviewer award at the conference COLT'10
Apr. 2009	Invited professor at the Libera Università Internazionale degli Studi Sociali, Roma, Italy
May 2008	Marie-Jeanne Laurent-Duhamel prize[a] of the French Statistical Society
June 2005	Student paper award at the conference COLT'05
Aug. 2003	Student paper award at the conference COLT'03

[a] For the best PhD thesis in theoretical statistics defended during the academic years 2004–05, 2005–06, 2006–07

Contents

CHAPTER 1

Foundations of the prediction of arbitrary sequences

INTRODUCTION. We focus in this chapter on a generic formulation of the problem of sequential prediction, which highlights the meta-statistical point of view. Formally, observations $y_1, \dots, y_t$ are to be predicted sequentially; no assumption is made on their generating process. In particular, these observations are not considered the realization of some underlying stochastic process whose parameters should be estimated in order to provide accurate predictions. Put differently, the problem at hand is not of a statistical nature.

However, a finite number of base predictors (indexed by $j = 1, \dots, N$) are available; at each time instance t, when the outcome to be predicted is y_t, they form a prediction $f_{j,t}$. These predictors may rely on some stochastic modeling and use statistical methods. The aim is to aggregate sequentially their predictions $f_{j,t}$ so as to output a combined forecast $\widehat{y}_t$ that is as accurate as possible.

Because of this the considered framework is of a meta-statistical nature. This chapter surveys some of its fundamental results and connects them to some other results of the classical statistical setting. This presentation is followed by an overview of my mathematical contributions to the foundations of the theory of sequential prediction of arbitrary sequences.

Table of contents

1.1 Two (related) settings of sequential prediction of arbitrary sequences

We start by describing the common features of the two settings. Two other features will make them different: whether the set of predictions $\mathcal{X}$ is convex or not, whether the observations may depend on the forecaster's predictions or not. A brief history of the field will then be addressed.

1.1.1 Common features of the two settings

The aim is to predict sequentially observations $y_1, y_2, \ldots$ lying in a given set $\mathcal{Y}$, on which no assumption or restriction is made. Unlike the classical framework of statistics, this sequence of observations is not to be modeled as the realization of a given underlying stochastic process. The problem at hand is therefore not to estimate the characteristics of such a process in order to predict its behavior and to form forecasts that are as accurate as possible.

Sets of observations and predictions. At each time instance t, the statistician is to output a forecast $\widehat{y}_t$ based on the past observations $y_1, \ldots, y_{t-1}$; it belongs to a set $\mathcal{X}$, possibly different from $\mathcal{Y}$. A typical case is when $\mathcal{X}$ is the convex hull of $\mathcal{Y}$: the occurrence of an event is to be predicted, i.e., $\mathcal{Y} = \{0, 1\}$, and to this end the statistician may output a probability of realization, i.e., he chooses an element of $\mathcal{X} = [0, 1]$.

Assessment of the quality of the predictions. The prediction $\widehat{y}_t$ is then compared to the observation y_t via a loss function $\ell : \mathcal{X} \times \mathcal{Y} \longrightarrow \mathbb{R}$, which is often non negative. The cumulative loss of the statistician on the first T time instances is then defined as

$$\sum_{t=1}^{T} \ell(\widehat{y}_t, y_t)$$

and his goal is to ensure that this loss is as small as possible.

Use of experts. A key ingredient is needed for this problem of prediction deprived of any stochastic assumption to be meaningful: the statistician may resort to some experts. The latter correspond to base predictors that output at each time instance a forecast based on the past observations.

More precisely, they come in finite number N and are indexed by $j = 1, \ldots, N$ (or by i when another ghost variable is required). Expert j provides at time instance t a forecast denoted by $f_{j,t} \in \mathcal{X}$ and that depends on $y_1, \ldots, y_{t-1}$ and possibly on other private information. The statistician can then form a combined forecast based not only on the past observations $y_1, \ldots, y_{t-1}$ but also on the past and present forecasts of the experts, $f_{j,s}$ where $1 \leqslant 1 \leqslant t$ and $j = 1, \ldots, N$. The consideration of the past forecasts of the experts is useful to assess in some sense the expected quality of their present forecasts.

Aim and methodology. The ultimate goal is to ensure that the cumulative loss of the statistician is small. A path to this will be to guarantee that this loss is not much larger than, e.g., the cumulative loss of the best expert,

$$\min_{j=1,\ldots,N} \sum_{t=1}^{T} \ell(f_{j,t}, y_t)\,,$$

where we however note that the index $j_T^\star$ of the expert achieving the minimum above may change over time and is in general not known in advance.

Two underlying assumptions to determine

The general description above is ambiguous (on purpose). Two features need to be detailed: whether the set of observations $\mathcal{X}$ is convex or not; whether the generating process of the observations y_t and the experts may react or not to the sequential forecasts of the statistician. Therefore, four instances of this general framework could be considered but for the sake of simplicity we only study the following two: on the one hand, the setting where $\mathcal{X}$ is convex and no reaction to the forecasts is allowed; and on the other hand, the one where $\mathcal{X}$ is arbitrary and where reactions are allowed.

1.1.2 First setting: Sequential convex aggregation

Here, $\mathcal{X}$ is convex and the sequence of observations $y_1, y_2, \dots$ is thought of as being fixed in advance but revealed element by element at each time instance t. The statistician is also constrained to only form forecasts $\widehat{y}_t$ obtained as convex combinations of the experts forecasts $f_{j,t}$.

Experts and nature. The generating process is independent of the statistician and is therefore identified to nature. As for the terminology of "experts", it stems from the fact that in addition to possibly relying on statistical techniques, they may also resort to contextual information, use numerical resources, and even call for human expertise. They will essentially be considered as black boxes in the rest of this chapter. Later on, when applications to real data will be discussed in Chapter 3, we will of course indicate for each application how the experts were constructed in practice.

Definition of a strategy of prediction by convex aggregation. Such a strategy $\mathcal{S}$ associates with the information available at the beginning of each time instance t (that is, to the past observations and to the present and past forecasts of the experts) a convex weight vector $\boldsymbol{p}_t = (p_{1,t}, \dots, p_{N,t})$ to be used to aggregate the experts forecasts in $\mathcal{X}$ as follows:

$$\widehat{y}_t = \sum_{j=1}^{N} p_{j,t} f_{j,t} \,.$$

Formally, the weight vectors $\boldsymbol{p}_t$ are chosen in the simplex $\mathcal{P}$ of $\mathbb{R}^N$, that is, they satisfy

$$\forall i \in \{1, \dots, N\}, \quad p_{i,t} \geqslant 0 \qquad \text{and} \qquad \sum_{j=1}^{N} p_{j,t} = 1 \,.$$

Figure 1.1 summarizes the prediction protocol considered in this setting.

Parameters: $\mathcal{Y}$, an arbitrary set of observations; $\mathcal{X}$, a convex set of predictions; N experts
Initialization: nature chooses a sequence of observations $y_1, y_2, \ldots$ in $\mathcal{Y}$
At each time instance $t = 1, 2, \ldots$,

1. The experts publicly output forecasts $f_{j,t} \in \mathcal{X}$, for $j = 1, \ldots, N$, based on the observations $y_1, \ldots, y_{t-1}$ and possibly on some own contextual information;
2. The statistician picks a convex weight vector $\boldsymbol{p}_t \in \mathcal{P}$ and forms the aggregated forecast in $\mathcal{X}$
$$\widehat{y}_t = \sum_{j=1}^{N} p_{j,t} f_{j,t} \,;$$
3. Nature reveals the observation $y_t \in \mathcal{Y}$.

Figure 1.1. The prediction protocol for the setting of sequential convex aggregation.

Assessment of the quality of a strategy via its regret

A strategy $\mathcal{S}$ cannot be assessed in an absolute way: when all experts are poor, no strategy is likely to exhibit a good performance. This is why a relative criterion –called the regret of a strategy $\mathcal{S}$– is usually considered; it compares the performance of $\mathcal{S}$ to the one of the best constant convex combination of the experts forecasts.

Notion of regret. We define the cumulative losses of $\mathcal{S}$ and of each convex weight vector $\boldsymbol{q} \in \mathcal{P}$ as, respectively,

$$\widehat{L}_T(\mathcal{S}) = \sum_{t=1}^{T} \ell(\widehat{y}_t, y_t) = \sum_{t=1}^{T} \ell\left(\sum_{j=1}^{N} p_{j,t} f_{j,t},\, y_t\right)$$
$$\text{and} \qquad L_T(\boldsymbol{q}) = \sum_{t=1}^{T} \ell\left(\sum_{j=1}^{N} q_j f_{j,t},\, y_t\right) .$$

The (convex) regret of $\mathcal{S}$ on the first T time instances is then defined as the difference between these cumulative losses,

$$R_T(\mathcal{S}) = \widehat{L}_T(\mathcal{S}) - \inf_{\boldsymbol{q} \in \mathcal{P}} L_T(\boldsymbol{q}) \,.$$

The quantities $\widehat{L}_T(\mathcal{S})$, $L_T(\boldsymbol{q})$, and $R_T(\mathcal{S})$ depend of course on the observations $y_1, \ldots, y_T$ and on the predictions of the experts even if, for the sake of simplicity, we do not explicitly recall this dependency in the notation.

Upper bounds on the regret. The regret $R_T(\mathcal{S})$ is at most of the order of T whenever the loss function is bounded. We aim at constructing strategies with vanishing per-round regrets, i.e., with $o(T)$ regrets; we in fact aim at the more ambitious goal of obtaining

uniform sublinear bounds on the regrets, where uniformity is over all sequences of observations and experts forecasts. The latter uniformity indicates that all possible sequences of observations in $\mathcal{Y}$ can happen; it gives rise to the terminology of arbitrary (or individual) sequences on the one hand and of robust aggregation on the other hand.

Aim 1.1. Design strategies $\mathcal{S}$ of prediction by convex aggregation minimizing the regret, that is, such that

$$\limsup_{T\to\infty} \sup \left\{ \frac{R_T(\mathcal{S})}{T} \right\} \leqslant 0\,,$$

where the supremum is over all possible sequences of observations and of experts forecasts.

Interpretation as a meta-statistical problem

Trade-off between two errors: approximation versus estimation. The aim stated above refers to the minimization of the regret while the ultimate goal is to ensure that the cumulative loss of the statistician is small. But the decomposition

$$\widehat{L}_T(\mathcal{S}) = \inf_{\boldsymbol{q}\in\mathcal{P}} \{L_T(\boldsymbol{q})\} + R_T(\mathcal{S})$$

indicates that this cumulative loss is the sum of an approximation error (given by the cumulative loss of the best constant convex combination of the experts) and an estimation error (given by the regret, which measures the difficulty induced by the sequential constraint to come close to the performance of this best constant convex combination). We recall in passing that the value of the optimal constant convex weight vector over the time instances from 1 to T may vary with T.

In practice there exists a dilemma between the use of a sufficiently large number N of experts with varied enough behaviors (to ensure a small enough approximation error) and the fact that the regret $R_T(\mathcal{S})$ (the estimation error in some sense) of course increases with N. However we will see that in general this increase is mild enough: of the order of $\sqrt{\ln N}$. It thus seems that in practice the use of quite large a number of experts is beneficial.

How to cook up experts? The main question at hand is then to design the experts. For the time being we only identified each expert with a forecasting black box. We will explain in detail on the data sets studied in Chapter 3 how the experts were constructed but allude now to a generic mechanism to produce experts; the latter illustrates why the problem of sequential aggregation is of a meta-statistical nature.

In a classical statistical problem where the observations (y_t) are the realizations of some underlying stochastic process (Y_t), stochastic methods lead to random predictions: we denote by $f_{j,t}$ the realization of the forecast of the j–th stochastic method at time instance t. Put differently, we identify each method with an expert. Instead of selecting a given method we consider several of them and aggregate their predictions. This

aggregation is performed in a robust way not taking into account the possible stochastic nature of the observations. The advantage is that the stochastic methods usually crucially rely on one or more user parameter(s); the considered methodology is to create several instances of them, each with different sets of user parameters, which makes the precise tuning of these parameters less crucial.

In a nutshell. We considered in this section the robust and non-stochastic aggregation of base forecasters (the experts) that may however depend on stochastic methods. In this sense the problem at hand is of a meta-statistical nature: we do not aim at improving the individual performance of the predictors but at combining well their forecasts.

1.1.3 Second setting: Randomized prediction

When the prediction set $\mathcal{X}$ is not convex, it is not always possible or easy to form a legal aggregated forecast based on the experts forecasts. A simple way out of it to allow prediction strategies to pick an expert and follow its forecast. The following counter-example proves that it is necessary in general to pick the expert at random: let $\mathcal{X} = \mathcal{Y} = \{0, 1\}$, consider a loss function ℓ given by a distance on $\{0, 1\}^2$ and two experts with constant forecasts over time, 0 and 1 respectively. Because of this random choice, the forecasts of the statistician themselves are random, hence the terminology of "randomized prediction."

Another modification of the previous setting: reactions to the forecasts. We also assume now that the generating process of the outcomes may react to the forecasts output by the statistician. The statistician is playing against an adversary that has also a strategy. One can even have this adversary control the experts and choose their forecasts.

Random draws of experts forecasts. The statistician still chooses in this setting an element $\boldsymbol{p}_t = (p_{1,t}, \ldots, p_{N,t})$ in $\mathcal{P}$, but the latter is now interpreted as a true probability distribution. The index of an expert is then drawn at random according to $\boldsymbol{p}_t$ and the statistician simply outputs the forecast of this expert. The corresponding prediction protocol is summarized in Figure 1.2.

Hidden random dependencies. We denote by σ and τ the respective strategies of the statistician and of the adversary. We do not define them formally in this general framework yet but will do it in a simplified randomized setting in Section 2.1. The last item of the description in Figure 1.2 indicates however (though in an informal way) that they associate with some past information their choices for the present time instance. For now we underline some difficulties thanks to an example. For instance, the choice of y_t at a given time instance $t \geqslant 2$ depends, among others, on $\widehat{y}_1, \ldots, \widehat{y}_{t-1}$, hence on the random variables $I_1, \ldots, I_{t-1}$. Thus, even when the strategy τ is deterministic, the resulting observations y_t are random variables; we call a deterministic strategy any strategy that associates with the available information a deterministic element

Parameters: $\mathcal{Y}$, an arbitrary set of observations; $\mathcal{X}$, an arbitrary set of predictions; N experts

At each time instance $t = 1, 2, \ldots,$

1. Based on the information provided by past time instances, the adversary picks an observation $y_t \in \mathcal{Y}$ and the forecasts $f_{j,t} \in \mathcal{X}$ of the experts $j = 1, \ldots, N$;
2. Only the experts forecasts are revealed to the statistician for now;
3. Based on the latter and on the information provided by past time instances, the statistician chooses a probability distribution $\boldsymbol{p}_t \in \mathcal{P}$, draws an expert index I_t at random according to $\boldsymbol{p}_t$, and outputs the forecast
$$\widehat{y}_t = f_{I_t,t}\,;$$
4. The adversary and the statistician publicly reveal their choices, that is, the observation $y_t \in \mathcal{Y}$ and the forecast $\widehat{y}_t \in \mathcal{X}$ (as well as the probability distribution $\boldsymbol{p}_t$ and the index I_t); both of them will recall these quantities in the next rounds and will be able to base their decisions on them.

Figure 1.2. The prediction protocol for the setting of randomized prediction.

of $\mathcal{X}^N \times \mathcal{Y}$ in the first item of the description in Figure 1.2. This argument extends to all quantities at hand, namely, to the experts forecasts $f_{j,t}$ and to the probability distributions $\boldsymbol{p}_t$ used for the random draws of experts indexes.

Extension of the definition of the regret

Here again –and for the same reasons as above– the evaluation of a strategy σ cannot be carried out in an absolute manner: if the adversary only picks poor experts forecasts no randomized prediction strategy is likely to perform well.

Regret with respect to the best expert, all things being equal. The cumulative losses of the statistician and of each expert j depend on the strategies σ and τ they respectively use: they are defined as

$$\widehat{L}_T(\sigma,\tau) = \sum_{t=1}^{T} \ell(\widehat{y}_t, y_t) = \sum_{t=1}^{T} \ell(f_{I_t,t}, y_t)$$

$$\text{and} \qquad L_{j,T}(\sigma,\tau) = \sum_{t=1}^{T} \ell(f_{j,t}, y_t)\,.$$

The regret of σ against τ on the first T time instances is then given by the difference between these cumulative losses,

$$R_T(\sigma,\tau) = \widehat{L}_T(\sigma,\tau) - \min_{j=1,\ldots,N} L_{j,T}(\sigma,\tau)\,. \tag{1.1}$$

Upper bounds on the regret. The regret $R_T(\sigma,\tau)$ is at most of the order of T when the loss function is bounded. Here again, we aim at constructing strategies σ with a vanishing per-round regret, no matter what the strategy τ of the adversary is. However, it is not possible in general to ensure a uniform bound on the regret, where uniformity would be with respect to all possible strategies τ.

Aim 1.2. Design randomized prediction strategies σ minimizing the regret, that is, such that

$$\sup_{\tau}\left\{\limsup_{T\to\infty}\frac{R_T(\sigma,\tau)}{T}\right\}\leqslant 0 \qquad \text{a.s.},$$

where the supremum is over all possible strategies of the adversary (and where the almost-sure qualification is with respect to the auxiliary randomizations used by the statistician and possibly by his adversary).

A setting with a somewhat uneasy interpretation

The issue: what the regret does not measure. We underline that the comparison to the best expert is performed in hindsight all things being equal, which raises an interpretation issue in the present setting where the adversary reacts to the forecasts of the statistician. Indeed, for a given expert j, if the statistician has consistently output the forecast of this expert, he would not have got in general $L_{j,T}(\sigma,\tau)$ as his cumulative loss –but rather $L_{j,T}(\sigma^j,\tau)$, where σ^j is the notation for the strategy that picks $\boldsymbol{p}_t=\delta_j$, the Dirac mass on j, independently of the past and of the index of the time instance. In particular, one would ideally wish to bound the difference between

$$\widehat{L}_T(\sigma,\tau) \qquad \text{and} \qquad \min_{j=1,\dots,N} L_{j,T}(\sigma^j,\tau)\,;$$

but this is –in general– not feasible and this is why the definition of the regret according to $R_T(\sigma,\tau)$ is considered.

Solutions. On the one hand [dFM03] suggests to restrict the set of possible strategies τ of the adversary to a subclass formed by the strategies with bounded rationality. On the other hand, if one is not ready to perform this restriction –as is my case– the defense of the notion of regret $R_T(\sigma,\tau)$ will not be intrinsic and elementary anymore, as was the case in the setting of sequential convex aggregation; several pages at the beginning of Chapter 2 will therefore be devoted to such a defense by means of convergences to sets of equilibria in the context of repeated games.

1.1.4 A brief history of the sequential prediction of non-stochastic sequences

The first contributions to the sequential prediction of arbitrary sequences (chosen or not by an adversary) were published in the 50s by Hannan [Han57] and Blackwell [Bla56], two statisticians who stated fundamental results for the theory of repeated games in these

two articles. Cover [Cov65] proposed the first study of the minimax orders of magnitude of the regret, in the context of the prediction of binary sequences A related setting is the compression of arbitrary sequences in information theory, where the pioneering results were written by Ziv [Ziv78, Ziv80] and Lempel and Ziv [LZ76, ZL77]; they showed how to compress an arbitrary data sequence as well as the best finite automaton. Finally, the introduction of the problem in machine learning was performed by Littlestone and Warmuth [LW94] and Vovk [Vov90]; Cesa-Bianchi, Freund, Haussler, Helmbold, Schapire and Warmuth [CBFH$^+$97], Foster [Fos91], Freund and Schapire [FS97], and Vovk [Vov98] conceived some of the most fundamental results in the field. A thorough and detailed state of the art as well as an historical survey of the advances between the 50s and 2006 can be found in the monography [CBL06].

1.2 Regret minimization with exponentially weighted averages

Never put all the weight on the current best expert! A natural strategy –but that fails to achieve the aims stated above as its regret can be of the order of T– is to use at time instance t the forecast of the expert that turned out to be best on the instances 1 to $t-1$. The issue with this strategy is essentially that it may assign far away weights, namely, 0 and 1, to two experts with very close cumulative performance. A wiser idea is to assign weights (or probabilities) $p_{j,t}$ which depend in a smoother way on the cumulative performance of expert j on the instances $1, \ldots, t-1$; the better the performance (the smaller the cumulative loss), the larger the weight –but as a precaution no weight should be null.

1.2.1 A fundamental result in a generic setting

The lemma stated below is one of the most fundamental –and also one of the most celebrated– results in prediction of individual sequences.

A generic setting. It is stated in a generic, non strategic, setting where only fixed-in-advance sequences of instantaneous losses are considered; we study therein bounds on a pseudo-regret defined in terms of convex weight vectors sequentially constructed based on the past instantaneous losses only. We will explain later on how to instantiate the results presented here to derive regret-minimizing strategies for both the settings of sequential convex aggregation and of randomized prediction (see Sections 1.2.2 and 1.2.3).

References. Several versions of this lemma were successively given by [LW94, Vov90, Vov98, CBFH$^+$97, FS97]. We reproduce below an elementary proof suggested by [CB99] and that can also be found in [CBL06, Section 2.2]. The strategy defined in the following lemma is called the exponentially weighted average strategy (with learning rate $\eta > 0$).

Lemma 1.3. Fix two real numbers $m \leqslant M$. For all $\eta > 0$ and for all arbitrary sequences of elements $\ell_{j,t} \in [m, M]$, where $j \in \{1, \dots, N\}$, and $t \in \{1, \dots, T\}$,

$$\sum_{t=1}^{T} \sum_{j=1}^{N} \mu_{j,t} \ell_{j,t} - \min_{i=1,\dots,N} \sum_{t=1}^{T} \ell_{i,t} \leqslant \frac{\ln N}{\eta} + \eta \frac{(M-m)^2}{8} T \,, \tag{1.2}$$

where for all $j = 1, \dots, N$, we let $\mu_{j,1} = 1/N$ and for $t \geqslant 2$,

$$\mu_{j,t} = \frac{\exp\left(-\eta \sum_{s=1}^{t-1} \ell_{j,s}\right)}{\sum_{i=1}^{N} \exp\left(-\eta \sum_{s=1}^{t-1} \ell_{i,s}\right)} \,. \tag{1.3}$$

Proof. The proof is based on Hoeffding's lemma, which we recall first. Consider a bounded random variable X with values in $[m, M]$; then, for all $s \in \mathbb{R}$,

$$\ln \mathbb{E}\left[e^{sX}\right] \leqslant s\, \mathbb{E}[X] + \frac{s^2}{8}(M-m)^2 \,. \tag{1.4}$$

In particular, for all $t = 1, 2, \dots$, with the convention (for the case where $t = 1$) that a sum over no element is null,

$$-\eta \sum_{j=1}^{N} \mu_{j,t} \ell_{j,t} \geqslant \ln \frac{\sum_{j=1}^{N} \exp\left(-\eta \sum_{s=1}^{t} \ell_{j,s}\right)}{\sum_{i=1}^{N} \exp\left(-\eta \sum_{s=1}^{t-1} \ell_{i,s}\right)} - \frac{\eta^2}{8}(M-m)^2 \,;$$

summing these inequalities over t and dividing both sides by $-\eta < 0$ yield

$$\sum_{t=1}^{T} \sum_{j=1}^{N} \mu_{j,t} \ell_{j,t} \leqslant -\frac{1}{\eta} \ln \frac{\sum_{j=1}^{N} \exp\left(-\eta \sum_{s=1}^{T} \ell_{j,s}\right)}{N} + \eta \frac{(M-m)^2}{8} T \,.$$

The proof is concluded by lower bounding the sum of positive terms in the logarithm of the right-hand side by the largest of these terms. □

Optimal theoretical tuning of η when the number of instances T is fixed and known...

All regret bounds exhibited in the sequel will be of the form of the right-hand side of (1.2). It is thus crucial to check that the latter is indeed sublinear.

Optimization of the theoretical bound. When the number of instances T as well as the bounds m and M on the losses are known beforehand one can resort to the tuning $\eta = \big(1/(M-m)\big)\sqrt{(8 \ln N)/T}$, which minimizes the right-hand side of (1.2). The obtained uniform regret bound then equals $(M-m)\sqrt{(T/2) \ln N}$. But this tuning is barely feasible in practice: first, there is often no reason to know m and M beforehand; second and most importantly, T cannot be thought of as being fixed.

... But in fact $T \to \infty$

The aims 1.1 and 1.2 both require that the number of time instances T tends to infinity. But for all constant choices of $\eta > 0$, the right-hand side of (1.2) grows then linearly fast so that none of these aims can be fulfilled.

Automatic and sequential tuning of the learning rates. The trick to solve the issues above is to have the learning rate η depend on the past. The weights $\mu_t \in \mathcal{P}$ are now defined (component-wise) in the following way: for all experts j, we choose $\mu_{j,1} = 1/N$ and for $t \geqslant 2$,

$$\mu_{j,t} = \frac{\exp\left(-\eta_t \sum_{s=1}^{t-1} \ell_{j,s}\right)}{\sum_{i=1}^{N} \exp\left(-\eta_t \sum_{s=1}^{t-1} \ell_{i,s}\right)}\,, \tag{1.5}$$

where the learning rate $\eta_t > 0$ at instance t may depend on the past elements $\ell_{i,s}$, where $s \in \{1, \dots, t-1\}$ and $i \in \{1, \dots, N\}$. In fact, η_t must even depend on these elements since, for instance, no a priori knowledge on the values m or M is available in general.

Existence of a suitable strategy. The key result of [ACBG02] and [5] is stated somewhat informally below but will be made formal later on, in Section 1.4.

Theorem 1.4. *There exists an explicit definition of each of the learning rates $\eta_t > 0$ based solely on the elements $\ell_{i,s}$, where $s \in \{1, \dots, t-1\}$ and $i \in \{1, \dots, N\}$, such that the strategy (1.5) ensures the following uniform bound. For all real numbers $m \leqslant M$, for all arbitrary sequences of elements $\ell_{j,t} \in [m, M]$, where $j \in \{1, \dots, N\}$ and $t \in \mathbb{N}^\star$, for all values of $T \in \mathbb{N}^\star$,*

$$\sum_{t=1}^{T} \sum_{j=1}^{N} \mu_{j,t} \ell_{j,t} - \min_{i=1,\dots,N} \sum_{t=1}^{T} \ell_{i,t} \leqslant 2(M-m)\sqrt{T \ln N} + 6(M-m)(1 + \ln N)\,.$$

> Remark in passing. We focus on upper bounds on the regret in the rest of this chapter. They will all be optimal in some sense as the upper bounds of Lemma 1.3 and of Theorem 1.4 are also optimal in some sense. Such an optimality statement will only be detailed in the case of label-efficient prediction of Section 1.3.1.

1.2.2 Application to randomized prediction

Assumption 1.5. We assume in this section that the loss function $\ell : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$ is bounded, with values in an interval denoted by $[m, M]$ (but that is not necessarily known beforehand).

Analysis when the parameters T, m, and M are known beforehand

In this case it suffices to consider a constant learning rate $\eta > 0$.

Statement of the strategy. The strategy $\mathcal{E}_\eta$ –called randomized exponentially weighted average strategy– resorts to the uniform distribution for $\boldsymbol{p}_1$, that is, $p_{j,1} = 1/N$ for $j = 1, \dots, N$; and for time instances $t \geqslant 2$, it chooses the probability distributions $\boldsymbol{p}_t$ defined component-wise as follows: for all $j = 1, \dots, N$,

$$p_{j,t} = \frac{\exp\left(-\eta \sum_{s=1}^{t-1} \ell(f_{j,s}, y_s)\right)}{\sum_{i=1}^{N} \exp\left(-\eta \sum_{s=1}^{t-1} \ell(f_{i,s}, y_s)\right)} \,. \tag{1.6}$$

It then draws an expert index I_t at random according to $\boldsymbol{p}_t$ and outputs the forecast $\widehat{y}_t = f_{I_t,t}$.

Analysis. By applying Lemma 1.3 (which holds in a deterministic manner for all sequences of real numbers $\ell_{j,t}$) to the random variables $\ell(f_{j,t}, y_t)$, one gets the almost-sure bound

$$\sum_{t=1}^{T} \sum_{j=1}^{N} p_{j,t} \ell(f_{j,t}, y_t) - \min_{i=1,\dots,N} \sum_{t=1}^{T} \ell(f_{i,t}, y_t) \leqslant \frac{\ln N}{\eta} + \eta \frac{(M-m)^2}{8} T = (M-m)\sqrt{\frac{T}{2} \ln N} \tag{1.7}$$

when η is properly tuned as a function of N, m, M, and T. We denote by $\mathbb{E}_t$ the conditional expectation with respect to the choices made by the statistician and the adversary in the past time instances 1 to $t-1$. Since this conditional expectation fixes the value of $\boldsymbol{p}_t$ but not the one of the random choice of I_t according to $\boldsymbol{p}_t$, we get

$$\mathbb{E}_t\Big[\ell(f_{I_t,t}, y_t)\Big] = \sum_{j=1}^{N} p_{j,t} \ell(f_{j,t}, y_t) \,.$$

> Remark in passing. When the adversary's strategy τ is deterministic, i.e., when he does not resort to an auxiliary randomization, the conditional expectation $\mathbb{E}_t$ is exactly the one with respect to the past indexes $I_1, \dots, I_{t-1}$.

The Hoeffding–Azuma inequality next ensures that with probability at least $1 - \delta$,

$$\sum_{t=1}^{T} \ell(f_{I_t,t}, y_t) - \sum_{t=1}^{T} \sum_{j=1}^{N} p_{j,t} \ell(f_{j,t}, y_t) \leqslant (M-m)\sqrt{\frac{T}{2} \ln \frac{1}{\delta}} \,. \tag{1.8}$$

Combining (1.7) and (1.8) shows that for all time instances T, there exists a tuning $\eta_T^\star$ for the learning rate (as a function of T, N, m, and M, even if only the dependency in T is made explicit in the notation) such that the regret of $\mathcal{E}_{\eta_T^\star}$ is bounded as follows. For all strategies τ of the adversary and with probability at least $1 - \delta$,

$$R_T(\mathcal{E}_{\eta_T^\star}, \tau) \leqslant (M-m)\sqrt{\frac{T}{2}} \left(\sqrt{\ln N} + \sqrt{\ln \frac{1}{\delta}}\right). \tag{1.9}$$

How to minimize the regret (how to fulfill Aim 1.2)

We denote by $\mathcal{E}_{\mathrm{adapt}}$ the strategy obtained by instantiating the strategy of Theorem 1.4 to the random variables $\ell(f_{j,t}, y_t)$, exactly as we derived above the strategies $\mathcal{E}_\eta$ from Lemma 1.3.

This theorem and the Hoeffding–Azuma inequality show that for all strategies τ of the adversary, all time instances T, and all confidence levels $1 - \delta_T \in]0, 1[$,

$$R_T(\mathcal{E}_{\mathrm{adapt}}, \tau) \leqslant (M-m)\sqrt{T}\left(2\sqrt{\ln N} + \sqrt{\frac{1}{2}\ln\frac{1}{\delta_T}}\right) + 6(M-m)(1+\ln N)\,.$$

In particular, taking $\delta_T = 1/T^2$, the Borel–Cantelli lemma implies that for all strategies τ of the adversary,

$$\limsup_{T\to\infty} \frac{R_T(\mathcal{E}_{\mathrm{adapt}}, \tau)}{(M-m)\sqrt{T\ln T}} \leqslant 1 \qquad \text{a.s.},$$

which shows among others that $\mathcal{E}_{\mathrm{adapt}}$ minimizes the regret (that it fulfills Aim 1.2).

> Remark in passing. By resorting above to a *maximal* version of the Hoeffding–Azuma inequality, by applying it at the instances of the form $T_r = 2^r$, and by combining it with the Borel–Cantelli lemma, one even gets that for all strategies τ of the adversary,
>
> $$\limsup_{T\to\infty} \frac{R_T(\mathcal{E}_{\mathrm{adapt}}, \tau)}{(M-m)\sqrt{2T\ln\ln T}} \leqslant 1 \qquad \text{a.s.},$$
>
> which resembles the law of the iterated logarithm (the latter states that the $\sqrt{\ln\ln T}$ term is necessary). This is to be compared with the results in the setting of convex aggregation, where the order of magnitude of the regret in T is $\sqrt{T}$, without any additional logarithmic factor in T.

1.2.3 Application to sequential convex aggregation

Assumption 1.6. We assume in this section that $\mathcal{X}$ is a bounded convex subset of $\mathbb{R}^d$ and that for all $y \in \mathcal{Y}$, the functions $\ell(\,\cdot\,, y)$ are convex and differentiable over $\mathcal{X}$, with gradients denoted by $\nabla\ell(\,\cdot\,, y)$; in addition, these gradients are uniformly bounded in the supremum norm as y varies.

Plain linearization is not sufficient. To apply Lemma 1.3 we need to upper bound the regret by a a quantity linear in the convex weight vectors $\boldsymbol{p}_t$. However the linear bound that proceeds from the convexity of ℓ in its first argument,

$$\sum_{t=1}^{T} \ell\left(\sum_{j=1}^{N} p_{j,t} f_{j,t},\, y_t\right) \leqslant \sum_{t=1}^{T}\sum_{j=1}^{N} p_{j,t}\,\ell(f_{j,t}, y_t)\,,$$

is too crude to allows a control of the cumulative loss of the statistician against smaller quantities than the cumulative loss of the best single expert.

Parameter: learning rate $\eta > 0$

Initialization : $\boldsymbol{p}_1$ is the uniform weight vector, that is, $p_{j,1} = 1/N$ for $j = 1, \dots, N$

For all time instances $t = 2, 3, \dots, T$, the convex weight vector $\boldsymbol{p}_t$ is defined component-wise as follows: for all $j = 1, \dots, N$,

$$p_{j,t} = \frac{\exp\left(-\eta \sum_{s=1}^{t-1} \widetilde{\ell}_{j,s}\right)}{\sum_{i=1}^{N} \exp\left(-\eta \sum_{s=1}^{t-1} \widetilde{\ell}_{i,s}\right)} \,,$$

where the pseudo-losses equal

$$\widetilde{\ell}_{j,s} = \nabla \ell \left(\sum_{i=1}^{N} p_{i,s} f_{i,s},\, y_s \right) \cdot f_{j,s} \,.$$

Figure 1.3. The strategy $\mathcal{E}_\eta^{\mathrm{grad}}$ uses an exponentially weighted average of the cumulative gradients of the losses.

Use of a slope inequality. What follows was proposed by [KW97, CB99] and can also be found in [CBL06, Section 2.5]. We stated an explicit condition of differentiability in Assumption 1.6 –despite the fact that all convex functions are subdifferentiable on the interior of the convex sets on which they are defined, a property that would (almost) have been sufficient here. In any case, these (sub)differentiability properties lead to so-called slope inequalities: for all convex weight vectors $\boldsymbol{p}$ et $\boldsymbol{q}$, for all forecasts $f_1, \dots, f_N \in \mathcal{X}$, and all observations $y \in \mathcal{Y}$,

$$\ell \left(\sum_{j=1}^{N} p_j f_j,\, y \right) - \ell \left(\sum_{j=1}^{N} q_j f_j,\, y \right) \leqslant \nabla \ell \left(\sum_{j=1}^{N} p_j f_j,\, y \right) \cdot \left(\sum_{j=1}^{N} p_j f_j - \sum_{j=1}^{N} q_j f_j \right) . \quad (1.10)$$

We next define the following pseudo-losses, for each expert $j \in \{1, \dots, N\}$ at each time instance $t \in \{1, \dots, T\}$:

$$\widetilde{\ell}_{j,t} = \nabla \ell \left(\sum_{i=1}^{N} p_{i,t} f_{i,t},\, y_t \right) \cdot f_{j,t} \quad (1.11)$$

and consider the family of aggregation strategies of Figure 1.3; each of them is parameterized by $\eta > 0$ and will be referred to as $\mathcal{E}_\eta^{\mathrm{grad}}$. The regret of $\mathcal{E}_\eta^{\mathrm{grad}}$ is then bounded by

$$\begin{aligned} R_T(\mathcal{E}_\eta^{\mathrm{grad}}) &= \sup_{\boldsymbol{q} \in \mathcal{P}} \sum_{t=1}^{T} \left(\ell \left(\sum_{j=1}^{N} p_{j,t} f_{j,t},\, y_t \right) - \ell \left(\sum_{j=1}^{N} q_j f_{j,t},\, y_t \right) \right) \\ &\leqslant \sup_{\boldsymbol{q} \in \mathcal{P}} \sum_{t=1}^{T} \left(\sum_{j=1}^{N} p_{j,t} \widetilde{\ell}_{j,t} - \sum_{j=1}^{N} q_j \widetilde{\ell}_{j,t} \right) = \sum_{t=1}^{T} \sum_{j=1}^{N} p_{j,t} \widetilde{\ell}_{j,t} - \min_{i=1,\dots,N} \sum_{t=1}^{T} \widetilde{\ell}_{i,t} \,, \end{aligned}$$

where the inequality follows from (1.10) and the second equality from the fact that the obtained upper bound is linear in $\boldsymbol{q}$, hence is maximized by a weight vector equal to a Dirac mass on a given expert. A straightforward application of Lemma 1.3 leads to the main result of this section.

Theorem 1.7. *When Assumption 1.6 is satisfied, the pseudo-losses defined in (1.11) are bounded, with values in some interval $[-C, C]$, and for all $\eta > 0$,*

$$\sup\left\{R_T(\mathcal{E}_\eta^{\mathrm{grad}})\right\} = \sup\left\{\widehat{L}_T(\mathcal{E}_\eta^{\mathrm{grad}}) - \inf_{\boldsymbol{q}\in\mathcal{P}} L_T(\boldsymbol{q})\right\} \leqslant \frac{\ln N}{\eta} + \eta\frac{C^2}{2}T\,,$$

where the supremum is over all possible sequences of observations and of experts forecasts. In particular, the tuning $\eta^\star = (1/C)\sqrt{(2\ln N)/T}$ leads to the upper bound

$$\sup\left\{R_T(\mathcal{E}_{\eta^\star})\right\} \leqslant C\sqrt{2T\ln N}\,.$$

Calibration and Aim 1.1. Here again, a calibration issue is raised for the tuning of η, whose theoretical optimal value in the theorem above depends on C and T; C is the bound on the range of the values of the pseudo-losses and is possibly unknown beforehand, while the number of time instances T is to tend to infinity. The same patch as in the previous section can be applied, namely, instantiating the strategy of Theorem 1.4 on the pseudo-losses (1.11) instead of resorting to Lemma 1.3. (We note that of course the values of the pseudo-losses strongly depend on the strategy used.) Aim 1.1 is then fulfilled with a uniform upper bound on the regret of the order of $C\sqrt{T\ln N}$.

> Remark in passing. Strong convexity assumptions (that follow, e.g., from a uniform upper bound on the gradients and a uniform lower bound on the eigenvalues of the Hessian matrices) yield sharper bounds on the regret, of the order (in T) of $\ln T$.

1.3 Contributions to randomized prediction [2, 12]

We present rather briefly two such contributions. Their common feature from a theoretical viewpoint is that they both rely on the use of exponentially weighted averages. The techniques used in the article [2] are similar to the ones needed later on in the articles [3, 6] discussed in the next chapter.

1.3.1 Label-efficient randomized prediction [2]

Presentation. This variation of the setting of plain randomized prediction was introduced by [HP97]. Here, querying the outcomes y_t after outputting his forecast is costly for the statistician. He has a limited budget to do so, the (dynamic) budget being modeled by a non-decreasing function $B : \mathbb{N}^\star \to \mathbb{N}^\star$ indicating that at each time instance $t \geqslant 1$, not more than $B(t)$ queries of y_t have been issued. The prediction protocol of Figure 1.2

is then modified as follows. The function B is added to the parameters known to the statistician and the item 4. of the iterations is replaced by the following two items:

> 4. The statistician reveals his choices $\widehat{y}_t$, I_t, and $\boldsymbol{p}_t$ to the adversary, who will recall them in the next rounds;
> 5. If the statistician has accessed less than $B(t)$ observations up to now, and only in this case, he may issue a query for $y_t \in \mathcal{Y}$, whose value he will then recall in the next rounds; this is the only way for him to compute his loss and the losses of the experts.

Aim. The target is still to fulfill Aim 1.2. This, of course, is only feasible when the function B grows quickly enough: for instance, when B is identically null, the statistician gets no feedback and no strategy can fulfill the prescribed aim.

A simple trick: estimate the unobserved quantities

Simplifying assumptions to have a gentle start. We start with a simplified framework where the number of time instances T is fixed and known, where the loss function ℓ takes its values in a known range of the form $[0, M]$, and where the budget function satisfies $B(1) = B(2) = \ldots = B(T)$, the common value of these budgets being denoted by B_T.

Estimators of the losses. We denote by $Z_1, \ldots, Z_T$ a sequence of independent random variables (they are also independent of all other considered random variables, e.g., of the other auxiliary randomizations used). Their common distribution is chosen to be a Bernoulli distribution with parameter $p \in]0,1[$; the latter will be defined (thanks to Bernstein's inequality) such that with a prescribed confidence level, $Z_1 + \ldots + Z_T \leqslant B_T$. Our strategy issues a query for y_t depending on the auxiliary randomization: it does so if and only if $Z_t = 1$ and the budget has not been overrun yet. For all time instances $t \geqslant 1$ and all expert indexes $j \in \{1, \ldots, N\}$, the following random variable is an estimator of $\ell(f_{j,t}, y_t)$:

$$\widehat{\ell}_{j,t} = \begin{cases} \dfrac{\ell(f_{j,t}, y_t)}{p} & \text{if } Z_t = 1 \text{ and } 1 + \displaystyle\sum_{s=1}^{t-1} Z_s \leqslant B_T; \\ 0 & \text{otherwise.} \end{cases}$$

Conditionally unbiased estimators. We denote by $\mathbb{E}_t$ the conditional expectation with respect to the information gathered by the statistician on the time instances 1 to $t-1$ (including the auxiliary randomizations $Z_1, \ldots, Z_{t-1}$) and with respect to the choices of the adversary at time instance t (i.e., with respect to the experts forecasts $f_{j,t}$ and the observation y_t). By construction,

$$\mathbb{E}_t\left[\widehat{\ell}_{j,t}\right] = \ell(f_{j,t}, y_t) \qquad \text{on } \left\{1 + Z_1 + \ldots + Z_{t-1} \leqslant B_T\right\}.$$

The exhibited estimators are therefore expected to perform well.

Substitution of these estimators for the losses in the weighted averages. The considered strategy picks the uniform probability distribution at $t = 1$ and in the subsequent time instances $t \geqslant 2$, it uses the probability distributions $\boldsymbol{p}_t$ with components defined by substituting the above-introduced estimators for the true losses in (1.6), that is,

$$p_{j,t} = \frac{\exp\left(-\eta \sum_{s=1}^{t-1} \widehat{\ell}_{j,s}\right)}{\sum_{i=1}^{N} \exp\left(-\eta \sum_{s=1}^{t-1} \widehat{\ell}_{i,s}\right)} ;$$

and as was already discussed, the auxiliary randomizations Z_t dictate the queries of the observations. The strategy thus defined depends on two parameters η and p; it is denoted by $€_{\eta,p}$ (the euro symbol is used because the strategy maintains a budget).

Reference. [ACBFS02] already exhibited conditionally unbiased estimators of unobserved losses in another setting of limited feedback: the multi-armed bandit problems of Chapter 4.

Analysis (under the simplifying assumptions). It relies quite crucially on the fact that losses are nonnegative (and are bounded by a known quantity M). Indeed, Lemma 1.3 is adapted as follows: the inequalities

$$\forall x \in \mathbb{R}_+, \qquad e^{-x} \leqslant 1 - x + \frac{x^2}{2} \qquad \text{and} \qquad \forall u > -1, \qquad \ln(1+u) \leqslant u$$

are used instead of Hoeffding's lemma to replace the right-hand side of (1.2) by

$$\sum_{t=1}^{T} \sum_{j=1}^{N} \mu_{j,t} \ell_{j,t} - \min_{i=1,\dots,N} \sum_{t=1}^{T} \ell_{i,t} \leqslant \frac{\ln N}{\eta} + \frac{\eta}{2} \sum_{t=1}^{T} \sum_{j=1}^{N} p_{j,t}\, \ell_{j,t}^2 \,. \tag{1.12}$$

Now, since this inequality is deterministic, it can be instantiated on the random quantities $\widehat{\ell}_{j,t}$ to provide an almost-sure bound. Some elementary concentration results (among others, the maximal version of Bernstein's inequality) and a proper tuning of $p^\star$ and $\eta^\star$ as functions of T, B_T, M, and δ then ensure that the following statement is true with probability at least $1 - \delta$ and for all strategies τ of the adversary:

$$\max_{t \leqslant T} R_t(€_{\eta^*,p^\star}, \tau) \leqslant 8MT \sqrt{\frac{\ln(4N/\delta)}{B_T}} \,. \tag{1.13}$$

Back to Aim 1.2

We use the so-called doubling trick. The corresponding strategy € is obtained from the base strategies described above by considering regimes indexed by integers $r \geqslant 1$, of lengths 2^r, and thus starting at time instances $T_r = 2^r - 1$, and by taking a fresh start of $€_{\eta_r,p_r}$ (where p_r and η_r are adequately chosen) at the beginning of each regime. The intermediate bounds (1.13), which are uniform over time, then lead to the main result of this section.

Theorem 1.8. *Whenever $B(T) \gg \ln T \, \ln\ln T$, there exists a strategy $\in$ (based on the doubling trick and which requires only the knowledge of the budget function B and of the range $[0, M]$ in which the loss function ℓ takes its values) ensuring that the regret-minimization Aim 1.2 is fulfilled and that at each time instance $T \geqslant 1$, no more than $B(T)$ queries have been issued since the beginning.*

Brief mention of an optimality result

Because all quantities at hand (cumulative losses, regret) are homogeneous in m and M, we fix their values in this subsection and choose, for simplicity, $m = 0$ and $M = 1$.

Upper bounds on the expectation of the regret. By integration over δ, the high-probability bounds on the regret (1.9) and (1.13) yield in particular uniform upper bounds on the expectation of the regret, where the uniformity is with respect to all strategies τ of the adversary; they are of the respective orders of $\sqrt{T \ln N}$ and $T\sqrt{(\ln N)/B_T}$. (The former order of magnitude would also follow directly from the combination of Theorem 1.4 with the tower rule.)

Optimality of the stated label-efficient bound. The plain setting of randomized prediction follows from the label-efficient one by taking $B_T = T$, from which the order of magnitude $\sqrt{T \ln N}$ is recovered. It thus suffices to focus on the optimality of the stated label-efficient bound. The following theorem shows that a simple and non strategic adversary may already force the regret of any forecasting strategy to be of the above-mentioned orders of magnitude. This adversary is denoted by $\tau(y_1^T)$, fixes beforehand a sequence $y_1^T = (y_1, \ldots, y_T)$ of observations, and picks constant experts, which output forecasts equal to, say, their indexes: $f_{j,t} = j$ for all time instances $t \geqslant 1$ and all $j \in \{1, \ldots, N\}$.

Theorem 1.9. *Consider the prediction and observation sets $\mathcal{X} = \mathbb{N}$ and $\mathcal{Y} = [0, 1]$. There exists a loss function ℓ with values in $[0, 1]$ such that for all $N \geqslant 2$ and all pairs (T, B_T) with $T \geqslant B_T \geqslant 15 \ln(N - 1)$, the regret of any strategy σ of the statistician bound to query at most B_T observations during the first T time instances satisfies*

$$\sup_{y_1^T \in \mathcal{Y}^T} \left\{ \mathbb{E}\Big[R_T\Big(\sigma, \tau(y_1^T)\Big)\Big] \right\} \geqslant \frac{T}{10}\sqrt{\frac{\ln N}{B_T}} \, .$$

Proof techniques. The main ingredient in [2] to prove the theorem above is Fano's lemma. A similar result based on a different proof technique (the use of lower bounds on the classification error of binary outcomes) is provided by [CBL06, Theorem 6.4].

References. In the case where $B_T = T$, the first (asymptotic) proof of the optimality of the orders of magnitude $\sqrt{T \ln N}$ was provided by [CBFH+97]; it is based on the central limit theorem and uses that the expectation of the maximum of N independent standard Gaussian random variables is equivalent to $\sqrt{\ln N}$. A non-asymptotic optimality result

based on Pinsker's inequality was also stated by [ACBFS02] in a different setting of limited feedback, namely, the multi-armed bandit problems of Chapter 4.

1.3.2 Simultaneous predictions (multi-task learning) [12]

A structured prediction problem. This section focuses on an instance of a so-called structured prediction problem. The latter problems correspond to settings where there are many experts that can however be compactly described by a small number of parameters: the class of experts is a structured class. For example, in the shortest path problem, each expert is identified with a path from a root node to a destination node; in the compound expert problem (also known as tracking the best expert) there are few base experts and each expert of the class is a meta-expert given by an infinite sequence of these base experts; see Section 3.5.1. Here, the structured problem at hand is to forecast simultaneous observations.

A simple but computationally costly strategy. In all the problems mentioned above a brute force strategy can be used: exponentially weighted averages over all elements of the structured experts class, whose cardinality N is usually large. The regret bound scales like $\sqrt{\ln N}$ as a function of N, which is often acceptable. However the computational cost of this naive strategy that allocates a weight to each expert is in general prohibitive as it scales linearly with N. Thus, the question reduces to finding an efficient implementation of the mentioned strategy; this efficient implementation can no longer consider each expert separately and needs in particular to group the experts into subclasses.

Description of the model

We deal with K prediction tasks indexed by $k \in \{1, \dots, K\}$. To each of the latter correspond a prediction set $\mathcal{X}^{(k)}$, an observation set $\mathcal{Y}^{(k)}$, and a loss function $\ell^{(k)} : \mathcal{X}^{(k)} \times \mathcal{Y}^{(k)} \to \mathbb{R}$. In addition N experts are available; each expert $j \in \{1, \dots, N\}$ outputs at each time instance $t \geqslant 1$ a forecast $f_{j,t}^{(k)}$ for each task k.

Prediction protocol (initial unconstrained version). The prediction protocol of Figure 1.2 is adapted to this setting as follows. At each time instance $t \geqslant 1$, the adversary chooses for each expert j a vector of forecasts $\boldsymbol{f}_{j,t}$ (these vectors are revealed to the statistician) and a vector of observations (that remains hidden for the time being):

$$\boldsymbol{f}_{j,t} = \left(f_{j,t}^{(1)}, \dots, f_{j,t}^{(K)}\right) \qquad \text{and} \qquad \boldsymbol{y}_t = \left(y_t^{(1)}, \dots, y_t^{(K)}\right).$$

The statistician draws at random an element of $\{1, \dots, N\}^K$, according to a probability distribution denoted by $\boldsymbol{p}_t$. We denote by

$$\boldsymbol{I}_t = \left(I_t^{(1)}, \dots, I_t^{(K)}\right)$$

the drawn element; it indicates for each task k the index $I_t^{(k)}$ of the expert whose forecast is to be followed. That is, the statistician outputs the vector of forecasts

$$\widehat{\boldsymbol{y}}_t = \left(f^{(1)}_{I_t^{(1)},t}, \ldots, f^{(K)}_{I_t^{(K)},t} \right). \tag{1.14}$$

The statistician and the adversary then publicly reveal all quantities introduced above and they will recall them in the next time instances.

Assessment of the quality of the predictions. The individual losses of all tasks lead to a global loss function ℓ via an evaluation function $\psi : \mathbb{R}^K \to \mathbb{R}$ in the following way. The quality of the prediction provided by a vector of forecasts $\boldsymbol{f}_{j,t}$ on the observations $\boldsymbol{y}_t$ is assessed by

$$\ell(\boldsymbol{f}_{j,t}, \boldsymbol{y}_t) = \psi\Big(\ell^{(1)}\Big(f^{(1)}_{j,t}, y^{(1)}_t\Big), \ \ldots, \ \ell^{(K)}\Big(f^{(K)}_{j,t}, y^{(K)}_t\Big) \Big).$$

Some examples of evaluation functions for which we have been able to provide an efficient implementation are the sum, the minimum, and the maximum of the individual losses; they are respectively defined by

$$\psi(x_1, \ldots, x_K) = \sum_{k=1}^{K} x_k \,, \qquad \psi(x_1, \ldots, x_K) = \min\{x_1, \ldots, x_K\} \,,$$
$$\text{and} \qquad \psi(x_1, \ldots, x_K) = \max\{x_1, \ldots, x_K\} \,.$$

For the sake of concision of the notation, the global loss at time instance t of each K–tuple $\boldsymbol{j} = (j_1, \ldots, j_K)$ in $\{1, \ldots, N\}^K$ (meaning that the forecast of expert j_k is followed in each task k) is denoted by

$$\ell_t(\boldsymbol{j}) = \psi\Big(\ell^{(1)}\Big(f^{(1)}_{j_1,t}, y^{(1)}_t\Big), \ \ldots, \ \ell^{(K)}\Big(f^{(K)}_{j_K,t}, y^{(K)}_t\Big) \Big);$$

the choices of the adversary are therefore put in the notation ℓ_t. For instance, with the shorthand notation (1.14), one gets

$$\ell(\widehat{\boldsymbol{y}}_t, \boldsymbol{y}_t) = \ell_t(\boldsymbol{I}_t) \,.$$

Addition of a constraint between the prediction tasks. For the time being the different tasks are K unrelated prediction problems. When the evaluation function is the sum of the individual losses it even suffices to run in parallel K base prediction strategies, one for each task; this yields a procedure with a computational complexity of the order of NK when, e.g., the base strategies are given by exponentially weighted average strategies.

Therefore we need to relate the tasks and our way out of it is to define a concept of legal predictions. Legal predictions are identified with the choice of a vector of experts

indexes in a strict subset $\mathcal{L}$ of $\{1,\dots,N\}^K$: at each time instance $t \geqslant 1$, the drawn indexes need to satisfy $\boldsymbol{I}_t \in \mathcal{L}$. In the definition of the regret the cumulative loss of the statistician will then also be compared to the best constant element in $\mathcal{L}$. As is illustrated below the definition and the structure of $\mathcal{L}$ model in some sense the links between the different prediction tasks at hand.

Adaptation of the notion of regret. The definition (1.1) is adapted as follows in the context of simultaneous predictions with constraints: the regret of a strategy σ of the statistician against a strategy τ of the adversary when being bound by the constraint $\mathcal{L}$ equals

$$R^{\mathrm{SP}}(\sigma,\tau) = \sum_{t=1}^{T} \ell_t(\boldsymbol{I}_t) - \min_{\boldsymbol{j}\in\mathcal{L}} \sum_{t=1}^{T} \ell_t(\boldsymbol{j})\,. \tag{1.15}$$

The aim is still to exhibit a (computationally efficient) strategy σ such that

$$\sup_{\tau} \left\{ \limsup_{T\to\infty} \frac{R^{\mathrm{SP}}(\sigma,\tau)}{T} \right\} \leqslant 0 \qquad \text{a.s.,} \tag{1.16}$$

where the supremum is over all strategies τ of the adversary.

Example of a set of constraints $\mathcal{L}$. We work out four examples in [12] but only reproduce one here (the easiest to describe), in which there is a cost for changing the expert from a task to the next one. A integer parameter $m \leqslant K-1$ is fixed and legal K–tuples $\boldsymbol{j}$ are those abiding by

$$\sum_{k=1}^{K-1} \mathbb{I}_{\{j_k \neq j_{k+1}\}} \leqslant m\,.$$

There are at least $(NK)^m/m!$ such K–tuples.

Comparison with previous work

There is no easy or natural definition of a multi-task setting for arbitrary sequences. Two models were proposed by [ABR07] and [DLS07].

The model of [ABR07]. At each time instance only one prediction task is to be performed; this task is chosen by the adversary and the statistician is not constrained in choosing his forecast. The relation betweens the tasks thus does not appear in the way the statistician forms his forecasts but only in the comparison class used in the definition of the regret. The latter is indeed taken as a strict subclass $\mathcal{L}'$ of the T–tuples of experts indexes, in a way similar to (1.15). However, since the statistician does not have to take into account the constraints of $\mathcal{L}'$ in making his own forecasts, it is not completely clear that the prediction tasks are sufficiently linked the ones to the others –parallel strategies could be employed (though they lead to suboptimal regret bounds).

The model of [DLS07]. It is formed by K (sub)problems of linear classification, where the losses are each measured by the hinge loss $x \mapsto (1-x)_+$ and are combined all together via an evaluation function ψ given by the Euclidian or the supremum norm over $\mathbb{R}^K$. The subproblems are only related through this global evaluation of the individual losses and through the comparison class in the definition of the regret, which is given by the consideration of the same linear hyperplane in all the (sub)problems.

Conclusion. The distinguishing feature of our model is that it limits the possible forecasts output by the statistician, who has to abide by the same constraints as the one used to form the comparison class in the definition of the regret.

Bound on the regret

Strategy. The exponentially weighted average strategy of Section 1.2.2 (tuned with a data-driven learning rate η_t as indicated in Theorem 1.4) forms at each time instance $t \geqslant 2$ the probability distribution $\boldsymbol{p}_t$ over $\mathcal{L}$ obtained as the following convex combination of Dirac masses $\delta_{\boldsymbol{j}}$:

$$\boldsymbol{p}_t = \sum_{\boldsymbol{j} \in \mathcal{L}} \frac{\exp\left(-\eta_t \sum_{s=1}^{t-1} \ell_s(\boldsymbol{j})\right)}{\sum_{\boldsymbol{i} \in \mathcal{L}} \exp\left(-\eta \sum_{s=1}^{t-1} \ell_s(\boldsymbol{i})\right)} \delta_{\boldsymbol{j}} \, ; \tag{1.17}$$

and for the first instance, $\boldsymbol{p}_1$ is the uniform distribution over $\mathcal{L}$. We denote by $\mathcal{M}$ the strategy thus obtained ($\mathcal{M}$ stands for multi-task learning).

Associated bound. We assume that the loss functions ℓ_t take uniformly bounded values, say in the range $[m, M]$. It then suffices to instantiate Théorème 1.4 as we already did in Section 1.2.2. For all strategies τ of the adversary and with probability at least $1-\delta$, the regret of $\mathcal{M}$ is less than

$$R_T(\mathcal{M}, \tau) \leqslant 2(M-m)\sqrt{T}\left(\sqrt{\ln|\mathcal{L}|} + \sqrt{\frac{1}{2}\ln\frac{1}{\delta}}\right) + 6(M-m)(1+\ln|\mathcal{L}|) \, ,$$

where $|\mathcal{L}|$ denotes the cardinality of $\mathcal{L}$; note that the log–cardinality $\ln|\mathcal{L}|$ is small (it is always smaller than $K \ln N$). A straightforward application of the Borel–Cantelli lemma next ensures that $\mathcal{M}$ achieves the prescribed aim (1.16).

Computationally efficient implementation in some cases

Random draws according to the $\boldsymbol{p}_t$ are sufficient. It only remains to see how to efficiently *draw* a vector of indexes $\boldsymbol{I}_t$ according to the distribution $\boldsymbol{p}_t$ in (1.17). This is of course to be performed without an explicit computation of $\boldsymbol{p}_t$ (the space complexity to store the values of the components of the latter would already be proportional to $\mathcal{L}$, hence be prohibitive!).

A hidden state space. We show in [12] how this computational cost depends on the structure of a hidden (inhomogeneous) Markov chain that can be set on $\mathcal{L}$, each element of the latter being seen as a realization of the first K states of this chain. We denote by S the cardinality of this hidden state space.

The random draw of $\boldsymbol{I}_t$ is then performed recursively: the last component $I_t^{(K)}$ is first drawn according to the last marginal of $\boldsymbol{p}_t$, then $I_t^{(K-1)}$ is drawn according to the $(K-1)$–th marginal of $\boldsymbol{p}_t$ conditionally to $I_t^{(K)}$, and so on. The main result is that to implement the described scheme it suffice to combine and update at most NKS quantities. The space complexity is thus proportional to NKS, while the computational complexity is slightly larger (by a multiplicative factor taking into account the number of possible transitions from a hidden space to another).

Back to the example (cost for changing the expert). Here, the space and computational complexities of the procedure described above are respectively of the orders of NKm and N^2Km. This is to be compared to the complexities of the order of $(NK)^m/m!$ suffered in the case of a direct computation and storage of (1.17).

1.4 Data-driven tuning of the parameters and data-dependent bounds [5]

A first desirable property: adaptive (data-driven) tuning of the parameters. This section is devoted, among others, to stating formally and providing elements of proof for Theorem 1.4. The latter is a fundamental result that was used to design fully automatic strategies minimizing the regret both in the settings of sequential convex aggregation and of randomized prediction. This feature of being fully automatic is obtained thanks to the sequence (η_t) of learning rates, which is constructed online based on the data (the past observations and experts forecasts). We detail this construction in this section and then state the regret bound: without knowing beforehand neither the number T of time rounds nor the range $[m, M]$ of the loss function (nor the bound C on the gradients of the losses), the statistician can guarantee that the regret is bounded by something of the order of $(M-m)\sqrt{T \ln N}$ (or $C\sqrt{T \ln N}$), which is the optimal order of magnitude in all parameters.

A second desirable property: sharper (data-dependent) bounds on the regret. The uniform regret bounds discussed in the previous sections are sometimes criticized for being too pessimistic: when one of the experts is much better than the other ones and has a small cumulative loss, the situation is simple enough to grasp and the regret should be much smaller than $\sqrt{T}$. One therefore aims at replacing the uniform bounds (that are valid for all sequences of losses) by data-dependent bounds (that depend strongly on the sequences of losses); of course, the worst-case values of the latter bound yield back the former bounds.

Aims of this section. We indicate how these two desirable properties can be met, based on [5] and on prior contributions. We will formulate all results to that end in the generic setting of Section 1.2.1 as we showed that all its results could be easily instantiated to the settings of convex aggregation and of randomized prediction. We recall that this generic setting consists of defining at each time instance $t \geqslant 1$ convex weight vectors μ_t over $\{1, \dots, N\}$ that only depend on the past losses $\ell_{j,s}$, where $s \leqslant t-1$ and $j \in \{1, \dots, N\}$, so as to upper bound the generic regret

$$\sum_{t=1}^{T} \sum_{j=1}^{N} \mu_{j,t} \ell_{j,t} - \min_{i=1,\dots,N} \sum_{t=1}^{T} \ell_{i,t} \,. \tag{1.18}$$

The bound we prove in this section may depend on the sequence of losses encountered, which can be assumed in the generic setting to be deterministic and fixed beforehand.

1.4.1 Review of the contributions prior to [5]

No tuning issue when polynomially weighted averages are used

The polynomially weighted average strategy essentially corresponds to the strategy exhibited in [Bla56] and was recently revisited by [CBL03]; it picks at time instances $t \geqslant 2$ the convex weight vectors μ_t defined component-wise by

$$\mu_{j,t} = \frac{\left(\sum_{t=1}^{T} \sum_{i=1}^{N} \mu_{i,t} \ell_{i,t} - \sum_{t=1}^{T} \ell_{j,t} \right)_+^{\alpha-1}}{\sum_{k=1}^{N} \left(\sum_{t=1}^{T} \sum_{i=1}^{N} \mu_{i,t} \ell_{i,t} - \sum_{t=1}^{T} \ell_{k,t} \right)_+^{\alpha-1}}$$

for all $j = 1, \dots, N$, where $(\,\cdot\,)_+$ denotes the nonnegative part of a real number and where the exponent satisfies $\alpha \geqslant 1$. The latter is the only parameter of this strategy. When the sequence of losses is bounded between m and M its generic regret is uniformly bounded by

$$(M-m)\sqrt{(\alpha-1)TN^{2/\alpha}} \leqslant (M-m)\sqrt{6T \ln N}$$

for the theoretical (almost) optimal choice $\alpha = 2 \ln N$; this tuning only depends on the number N of experts, a quantity always known beforehand, and not on possibly unknown parameters like m, M, and T as was the case for the simplest version of the exponentially weighted average strategy. The orders of magnitude of the bound on the generic regret of this strategy are furthermore optimal in all parameters. It is thus natural to wonder whether exponentially weighted average strategies (1.3) are worth the trouble.

Key point: exponentially weighted averages are useful when limited feedback only is available. Things get trickier for the polynomially weighted average strategies when the losses are not fully revealed at the end of a prediction round and when the thus unobserved losses need to be estimated as was the case in Section 1.3.1 and will be the case in

Section 2.2. It is true that [CBL06, Theorem 6.9] shows that polynomially weighted average strategies can have vanishing per-round regret in the context of the multi-armed bandit problems of Chapter 4 but this result is quite long and tedious to prove and is not associated with clear convergence rates of this per-round regret towards 0. This is to be compared to the simple arguments that led, e.g., to the upper bound (1.13) for an exponentially weighted average strategy.

Tuning of the parameters via periodic fresh starts

A folk solution to the tuning issues of the exponentially weighted average strategies (1.3) is called the doubling trick and is presented, e.g., in [CBL06, Section 2.3]. It consists of periodic fresh starts of this strategy with learning rates η_r that become smaller and smaller as the number r of already taken fresh starts increases. (Such a trick was used above in Theorem 1.8.) The learning rates η_r in each regime are taken of the form of the theoretical optimal value indicated by Lemma 1.3, that is, $\big(1/(M_r - m_r)\big)\sqrt{(8 \ln N)/2^r}$, where m_r and M_r denote estimates of the range of the losses computed in the $r-1$ past regimes. Each regime takes an end (and a new fresh start is taken) according to a stopping condition based on the value of T and/or linked to some severe violation of the estimated range $[m_r, M_r]$.

Key issue: loss of information and periodic lacks of efficiency. At each fresh start almost all information provided by past time instances is thrown out; only a small fraction of it is kept (the one summarized in the estimates m_r and M_r). In addition –and maybe most importantly– the statistician then uses again for a while convex weight vectors that are close the uniform weight vector, which is inefficient in practice.

Sharper regret bounds and partial calibration of the parameters for nonnegative losses

The analysis of the exponentially weighted average strategy (1.3) provided by [FS97] ensures the following data-dependent upper bound on its generic regret (1.18):

$$\sqrt{2ML_T^\star \ln N}\,, \qquad \text{where} \qquad L_T^\star = \min_{i=1,\dots,N} \sum_{t=1}^{T} \ell_{i,t}\,; \tag{1.19}$$

the losses need to take bounded values in some (known) range $[0, M]$ and the parameter η of the exponentially weighted average strategy is (illegally) tuned as a function of $L_T^\star$ (which cannot be known in advance) and of M. A legal tuning is however possible via a doubling trick but it still requires the knowledge of M. This bound is called the improvement for small losses.

The breakthrough: online tuning of the learning rates. [ACBG02] introduced the form (1.5) and provided an analysis of its performance. In particular it indicates, under the same

assumptions as in the previous paragraph (that is, knowing M and facing nonnegative losses), that choosing online learning rates η_t proportional to

$$\sqrt{\frac{\ln N}{M \sum_{s=1}^{t-1} \sum_{i=1}^{N} \mu_{i,s} \ell_{i,s}}} \tag{1.20}$$

implies the bound (1.19) up to essentially a multiplicative factor of 2 while no doubling trick or no beforehand knowledge of $L_T^\star$ is needed this time. The proof of this important result is sketched below.

What remains to be done. An online tuning not requiring the knowledge of the range of the losses needs to be developed; the assumption of nonnegativity of the losses must also be relaxed for the generic bounds to be instantiated in the setting of convex aggregation as we explain below.

The case of signed losses is crucial

[ANN04] was the first to consider the case where the losses $\ell_{j,t}$ are not necessarily nonnegative but are simply assumed to lie in a bounded range $[m, M]$, where $m \leqslant M$ are two real numbers. We do not assume that m and M are known; otherwise, the case of nonnegative losses would be recovered as soon as the strategies at hand would perform a straightforward translation of the losses.

Importance of this case. [ANN04] does not defend the interest of signed losses. To me it lies in the possibility of instantiating the generic regret bounds in the setting of convex aggregation via the pseudo-losses (1.11); the latter are indeed not necessarily nonnegative but they are usually bounded.

Target. It is thus desirable to deal with the online calibration of the parameters η_t when the bounds m and M are unknown and possibly negative numbers; ideally, we would like to recover bounds on the regret with the same orders of magnitude as when these parameters are known.

A remark: deviations around the generic regret bound in randomized prediction

In this section we only consider the generic regret but the bounds on it need to be instantiated to yield regret bounds, e.g., in the setting of randomized prediction. In the latter an additive term accounting for the high-probability deviations around the expected regret bound has to be considered. If the Hoeffding–Azuma inequality is applied as in (1.8), then the deviations are controlled by something of the order of $\sqrt{T \ln(1/\delta)}$. But this terms annihilates any data-dependent improvement exhibited on the generic regret: sharper concentration inequalities need to be used. For instance, Bernstein's

inequality for martingale difference sequences provides bounds on the likely deviations that are of the same order as the generic regret bound formed by the improvement for small losses stated above; for further details, see my PhD thesis [Sto05, pages 38–39].

1.4.2 Online calibration of the parameters with signed losses [5]

The core idea is to replace the denominator of (1.20) by a quantity that only depends on information collected in the past (and not on M); this quantity is expected to be homogeneous to square losses. This denominator actually came as the result of a first-order upper bound on a Laplace transform: we will replace this crude bound by a sharper second-order bound.

Modification of the proof of Lemma 1.3. [ACBG02] –see also the posterior simplifications in the proof provided by [CBL06, Section 2.3] and [GO07, Lemma 1]– states the following performance bound for the strategy (1.5): for all non-increasing sequences (η_t), possibly tuned online, the generic regret is bounded by

$$\sum_{t=1}^{T}\sum_{j=1}^{N}\mu_{j,t}\ell_{j,t} - \min_{i=1,\dots,N}\sum_{t=1}^{T}\ell_{i,t} \leqslant \frac{\ln N}{\eta_T} + \sum_{t=1}^{T}\Phi\big(\mu_t,\,(\ell_{j,t})_j,\,\eta_t\big)\,, \tag{1.21}$$

where the function Φ takes as arguments a convex weight vector μ, a loss vector $(\ell_1,\dots,\ell_N)$, and a learning rate η:

$$\Phi\big(\mu,\,(\ell_j)_j,\,\eta\big) = \frac{1}{\eta}\ln\left(\sum_{i=1}^{N}\mu_i\,e^{-\eta\left(\ell_i-\widehat{\ell}\right)}\right) \qquad \text{where} \quad \widehat{\ell} = \sum_{j=1}^{N}\mu_j\ell_j\,.$$

Assume that we already could upper bound the Φ terms by quantities of the form $\eta_t\,z_t$, with $z_t \geqslant 0$; that is, assume we face the following bound:

$$\sum_{t=1}^{T}\sum_{j=1}^{N}\mu_{j,t}\ell_{j,t} - \min_{i=1,\dots,N}\sum_{t=1}^{T}\ell_{i,t} \leqslant \frac{\ln N}{\eta_T} + \sum_{t=1}^{T}\eta_t z_t\,.$$

A natural online choice would then be given by

$$\eta_t = \sqrt{\frac{\ln N}{z_1+\ldots+z_{t-1}}}$$

for it yields the bound on the generic regret

$$\sum_{t=1}^{T}\sum_{j=1}^{N}\mu_{j,t}\ell_{j,t} - \min_{i=1,\dots,N}\sum_{t=1}^{T}\ell_{i,t} \leqslant 4\sqrt{\left(\sum_{t=1}^{T}z_t\right)\ln N}\,. \tag{1.22}$$

(The multiplicative constant 4 in front of the right-hand side can be improved.) The only question at hand is therefore to exhibit good z_t, that is, to sharply upper bound quantities of the form

$$\Psi_\eta(X) = \frac{1}{\eta^2}\,\ln\mathbb{E}\Big[e^{-\eta\left(X-\mathbb{E}[X]\right)}\Big]$$

for all $\eta > 0$ and all random variables X taking finitely many values.

Uniform bounds (zero-order bounds). The bound (1.2) in Lemma 1.3 corresponds to a constant sequence (η_t) and to the derivation of an upper bound on Ψ_η via Hoeffding's lemma –as indicated in (1.4). We want to improve on that.

Bounds for nonnegative losses (first-order bounds). We already noticed in (1.12) that for nonnegative losses bounded by M,

$$\Psi_\eta(X) \leqslant \frac{\eta}{2}\,\mathbb{E}\Big[X^2\Big] \leqslant \frac{\eta M}{2}\,\mathbb{E}[X]\,,$$

which corresponds (with the notation above) to

$$z_t = \frac{M}{2}\sum_{j=1}^N \mu_{j,t}\ell_{j,t}\,;$$

hence, by application of (1.22), the inequality

$$\sum_{t=1}^T\sum_{j=1}^N \mu_{j,t}\ell_{j,t} - \min_{i=1,\dots,N}\sum_{t=1}^T \ell_{i,t} \leqslant 2\sqrt{2}\,\sqrt{M\left(\sum_{t=1}^T\sum_{j=1}^N \mu_{j,t}\ell_{j,t}\right)\ln N}\,. \tag{1.23}$$

To get an improvement for small losses of the form (1.19) it then suffices to solve a second-order inequality; an extra multiplicative factor comes in front of the right-hand side of (1.19) and it measures the price for not knowing M and $L_T^\star$.

Second-order bounds with signed losses. The basic inequality $e^x \leqslant 1 + x + (e-2)\,x^2$ (valid for $x \leqslant 1$) yields that for all pairs (η, X) such that $\eta > 0$ and $\eta X \leqslant 1$ a.s.,

$$\Psi_\eta(X) \leqslant \frac{1}{\eta}\,\ln\Big(1 + (e-2)\eta^2\,\mathrm{Var}(X)\Big) \leqslant (e-2)\eta\,\mathrm{Var}(X)\,. \tag{1.24}$$

In particular, introducing for all t (respectively, T) a pseudo-variance v_t of the losses at time instance t (respectively, a cumulative pseudo-variance V_T of the losses up to time instance T),

$$v_t = \sum_{j=1}^N \mu_{j,t}\left(\ell_{j,t} - \sum_{i=1}^N \mu_{i,t}\ell_{i,t}\right)^2 \qquad \text{and} \qquad V_T = v_1 + \ldots + v_T\,,$$

we expect a bound on the generic regret of the form $\sqrt{V_T \ln N}$, via (1.22) and the choice $z_t = v_t$. However, a difficulty is that (1.24) requires a domination condition by 1, which is not always satisfied in practice; when it is not, one can still apply Hoeffding's bound (1.4). Based on these elements, we then show the following fundamental result, which is a more formal and more detailed statement of Theorem 1.4 above. We reproduce the result as it is stated in [5] even if [Ger10] recently realized that the multiplicative factor of 4 in the bound could be improved to

$$2\sqrt{(e-2)(\sqrt{2}-1)} \leqslant 2.64\,.$$

Theorem 1.10. *We consider the fully automatic strategy (1.5), where the learning rates are defined, for $t \geqslant 2$, as sole functions of the past losses according to*

$$\eta_t = \min\left\{ \frac{1}{E_{t-1}}, \gamma\sqrt{\frac{\ln N}{V_{t-1}}} \right\},$$

where $\gamma = \sqrt{2(\sqrt{2}-1)\,/\,(e-2)}$ and

$$E_{t-1} = \min\left\{ 2^k : \quad k \in \mathbb{Z} \quad \text{and} \quad \max_{s \leqslant t-1} \max_{i \neq j} |\ell_{i,s} - \ell_{j,s}| \leqslant 2^k \right\}.$$

Then, for all (possibly signed) real numbers $m \leqslant M$, for all arbitrary sequences of elements $\ell_{j,t} \in [m, M]$, where $j \in \{1, \dots, N\}$ and $t \in \mathbb{N}^\star$, for all $T \in \mathbb{N}^\star$,

$$\sum_{t=1}^{T} \sum_{j=1}^{N} \mu_{j,t} \ell_{j,t} - \min_{i=1,\dots,N} \sum_{t=1}^{T} \ell_{i,t} \leqslant 4\sqrt{V_T \ln N} + 6(M-m)(1+\ln N).$$

Corollaries and applications of this bound. In the theorem above the upper bound on the regret is in terms of V_T, which is not an intrinsic quantity but depends on the strategy. We explain here how to deal with that and recover bounds that only depend on the sequences of losses. A first straightforward idea is to note that the v_t are variance terms and hence are bounded by the squared half-range $(M-m)^2/4$. The substitution of the latter bound implies the bound initially stated in Theorem 1.4. On the other hand, since a variance is smaller than the expectation of the square quantity, one gets

$$V_T \leqslant \sum_{t=1}^{T} \sum_{j=1}^{N} \mu_{j,t} \ell_{j,t}^2,$$

which implies a bound on the regret similar to (1.23) when the losses are nonnegative, and hence an improvement for small losses. We actually show a stronger result in [5]: an improvement for small or large nonnegative losses.

Other comments. We underline that the regret bound of Theorem 1.10 is stable by translations of the losses, which is not the case of many regret bounds like, e.g., the improvement for small losses as it crucially relies on a nonnegativity assumption. In addition, as uncomfortable might seem to work with the $\sqrt{V_T}$ term, we stress that in the setting of randomized prediction it naturally comes into the picture when Bernstein's inequality for martingale difference sequences is used as a concentration argument in replacement of the Hoeffding–Azuma inequality in (1.8). This term is therefore unavoidable in some sense, the question being to determine whether there are even better ways to deal with it as the ones exposed in the previous paragraph.

1.5 Perspectives for future research

In this section the $\square$ symbols denote universal constants whose value is not computed explicitly and may even change from one occurrence of $\square$ to the next one.

Back to the cumulative loss...

We now stop considering the minimization of the regret as the final aim and get back to our initial wish: ensuring that the cumulative loss of the strategy is as small as possible.

Correspondance between the best approximation error and the best estimation error. The bounds on the cumulative losses that can be deduced from the regret bounds exhibited above are of the form

$$\sum_{t=1}^{T}\sum_{j=1}^{N}\mu_{j,t}\ell_{j,t} \leqslant \underbrace{\min_{i=1,\dots,N}\sum_{t=1}^{T}\ell_{i,t}}_{\text{approximation error}} + \underbrace{D\left(\min_{i=1,\dots,N}\sum_{t=1}^{T}\ell_{i,t}\right)}_{\text{estimation error}} \tag{1.25}$$

where the function D is either constant, equal to $\square\,(M-m)\sqrt{T\ln N}$ (in the case of uniform zero-order bounds), or is given by the mapping $x \mapsto \square\sqrt{x\ln N} + \square\, M\ln N$ (in the case of the improvement for small nonnegative losses).

The term being a function of D provides the regret bound. In the above-mentioned cases D is nondecreasing so that there is a correspondance between the expert with the smallest approximation error and the one with the smallest estimation error. There is no trade-off to perform between the two errors, since both are minimized by the same expert. This is not very realistic.

Desired form of the regret bound on the cumulative regret. This is why more general bounds of the form

$$\sum_{t=1}^{T}\sum_{j=1}^{N}\mu_{j,t}\ell_{j,t} \leqslant \min_{i=1,\dots,N}\left\{\sum_{t=1}^{T}\ell_{i,t} + D(\ell_{i,1},\dots,\ell_{i,T})\right\} \tag{1.26}$$

are desired, where the function $D:\mathbb{R}^T\to\mathbb{R}$ now has a more general form: it takes T values as arguments and though it is probably nondecreasing in each of its components, it is not necessarily a function of the sums of the losses.

A promising attempt... that however failed.

The initial motivation to [5] was to prove (1.26) with

$$D_{\text{SQ}}(x_1,\dots,x_T) = 2\sqrt{\sum_{t=1}^{T}x_t^2\ln N} + \square\,(M-m)\ln N$$

while [HK08] (whose aim was inspired by the open questions at the end of [5]) attempted to do so with

$$D_{\text{VAR}}(x_1,\dots,x_T) = \square\sqrt{\sum_{t=1}^{T}\left(x_t - \frac{1}{T}\sum_{s=1}^{T} x_s\right)^2 \ln N} + \square\,(M-m)\ln N\,.$$

Of course, both would have yielded upper bounds of the form (1.25) as special cases but could have brought substantial improvements over them.

A retrospective tuning would do the job. In [5] a new strategy called Prod_η is introduced; it is based on a parameter $\eta > 0$ but the convex weight vectors it prescribes do not depend solely on the cumulative losses (in particular, they are not given by exponentially or polynomially weighted averages of the cumulative losses). The generic regret of this strategy is bounded as follows: for all arbitrary sequences of elements $\ell_{j,t} \in [m,M]$, where $j \in \{1,\dots,N\}$ and $t \in \mathbb{N}^\star$, for all values of η such that $0 < \eta \leqslant 1/(2M)$, for all $T \in \mathbb{N}^\star$,

$$\sum_{t=1}^{T}\sum_{j=1}^{N} \mu_{j,t}\ell_{j,t} \leqslant \min_{i=1,\dots,N}\left\{\sum_{t=1}^{T}\ell_{i,t} + \frac{\ln N}{\eta} + \eta\sum_{t=1}^{T}\ell_{i,t}^2\right\}. \tag{1.27}$$

A proper retrospective tuning of η implies (1.26) with $D = D_{\text{SQ}}$: denoting

$$i_T^\star \in \operatorname*{arg\,min}_{i=1,\dots,N}\left\{\sum_{t=1}^{T}\ell_{i,t} + 2\sqrt{\sum_{t=1}^{T}\ell_{i,t}^2 \ln N}\right\},$$

it suffices to tune η as

$$\eta_T^\star = \sqrt{\frac{\ln N}{\sum_{t=1}^{T}\ell_{i_T^\star,t}^2}},$$

provided that the latter quantity is smaller than $1/(2M)$. One could think that routine online calibration techniques (as the ones described in the previous section) would yield almost the same bound in a true online fashion, i.e., with an online calibration of the learning rates. But a severe issue arose.

What we could prove (only). The online calibration techniques described above (the doubling trick or some online tuning of the learning parameters depending on the past losses) are based on quantities that increase with T (e.g., the cumulative loss of the strategy, the cumulative pseudo-variance, etc.). But the quantities at hand here, the

$$Q_T^\star = \sum_{t=1}^{T}\ell_{i_T^\star,t}^2\,,$$

are not necessarily increasing with T (because the values of $i_T^\star$ may change with T). One way around it is to consider the smallest non-decreasing upper bounds that are associated with them: $\max_{t \leqslant T} Q_t^\star$. This explains why our final bound on the generic regret of a fully automatic strategy based on the Prod_η strategies is of the form

$$\min_{i=1,\dots,N} \sum_{t=1}^{T} \ell_{i,t} + \square \sqrt{\max_{t \leqslant T} Q_t^\star \ln N} + \square \, (M-m) \ln N \, ;$$

the bound above is rather of the form (1.25) than of the desired form (1.26).

Similar issue for [HK08]. Even if the strategy described in the latter article can guarantee (1.26) for $D = D_{\text{VAR}}$ thanks to a suitable retrospective tuning (and based on exponentially weighted averages of the *penalized* cumulative losses), its online adaptation suffers of the same problem as described above for Prod_η.

Statement of the open problem

The question at hand is therefore either to prove inequalities of the form (1.26) or to show that no strategy can guarantee bounds of this kind. My intuition would rather be that they are achievable but a fundamental conceptual hurdle needs to be passed in terms of online calibration techniques.

CHAPTER 2

Interactions with the theory of repeated games

INTRODUCTION. This chapter focuses on a variant of the setting of randomized prediction studied in the previous chapter. Instead of a statistician facing an adversarial environment and trying to forecast its evolution, we now consider a couple of players each reacting to the other player's behavior. In addition rewards instead of losses will be received; each player will aim at maximizing the sum of obtained rewards. Finally, the prediction protocol is somewhat simplified: no expert is available and each player only has a finite number of actions at his disposal.

The underlying heuristic presented in this chapter is that whenever each player follows a strategy with good performance –in the sense that its regret (to be re-defined) is small– then some equilibrium situation is asymptotically reached. We will essentially discuss three such notions of equilibrium: first, in the case of a zero-sum game, the convergence of the mean payoffs of each player towards the value of the game; or, in the general case of more than two players, the convergence of the empirical frequencies of action profiles towards, second, the set of Hannan equilibria or towards, third, the set of correlated equilibria.

Table of contents

2.1 Definition and defense of the notion of regret

We actually will only state and study the problem in the case of two players; we do so for the sake of simplicity and indicate that all the results of this chapter (but the ones for zero-sum games) can be extend in a straightforward manner to the case of a finitely many players.

Notation. The two players are called players A and B and have finite action sets respectively denoted by $\mathcal{A} = \{1, \dots, N\}$ and $\mathcal{B} = \{1, \dots, M\}$ The payoff functions $\mathcal{A} \times \mathcal{B} \to \mathbb{R}$ are denoted by r for player A and s for player B. These functions r and s

are linearly extended on the simplexes $\Delta(\mathcal{A})$ and $\Delta(\mathcal{B})$ of probability distributions over $\mathcal{A}$ and $\mathcal{B}$: for all $\boldsymbol{p} = (p_i)_{i\in\mathcal{A}} \in \Delta(\mathcal{A})$ and $\boldsymbol{q} = (p_j)_{j\in\mathcal{B}} \in \Delta(\mathcal{B})$,

$$r(\boldsymbol{p},\boldsymbol{q}) = \sum_{i\in\mathcal{A}}\sum_{j\in\mathcal{B}} p_i\, q_j\, r(i,j) \qquad \text{and} \qquad s(\boldsymbol{p},\boldsymbol{q}) = \sum_{i\in\mathcal{A}}\sum_{j\in\mathcal{B}} p_i\, q_j\, s(i,j)\,.$$

Protocol of the repeated game. At each round $t = 1, 2, \ldots$, players A and B simultaneously pick their respective actions $I_t \in \mathcal{A}$ and $J_t \in \mathcal{B}$; the two actions are then revealed and the players obtain the respective payoffs $r(I_t, J_t)$ and $s(I_t, J_t)$. We call the pair (I_t, J_t) the action profile played at round t. The choices of I_t and J_t are made based on the past and thanks to an auxiliary randomization; that is, these actions are drawn at random according to probability distributions $\boldsymbol{p}_t$ and $\boldsymbol{q}_t$ over $\mathcal{A}$ and $\mathcal{B}$ that depend measurably on the history of action profiles $(I_1, J_1), \ldots, (I_{t-1}, J_{t-1})$ played in the past.

The strategies of A and B are given by sequences of mappings that for each $t \geqslant 1$, associate with all histories in $(\mathcal{A}\times\mathcal{B})^{t-1}$ an element respectively in $\Delta(\mathcal{A})$ or $\Delta(\mathcal{B})$; they will typically be denoted by σ and τ in the sequel. Many quantities to be introduced later on will depend on these strategies σ and τ but for the sake of simplicity, these dependencies will be omitted in the notation.

The empirical frequencies of the played actions are defined by

$$\overline{\boldsymbol{p}}_T = \frac{1}{T}\sum_{t=1}^{T} \delta_{I_t} \qquad \text{and} \qquad \overline{\boldsymbol{q}}_T = \frac{1}{T}\sum_{t=1}^{T} \delta_{J_t}\,.$$

Aims and quantities of interest. Each player aims at obtaining an (asymptotic) per-round payoff that is as large as possible; that is, the players A and B are respectively interested in the quantities

$$\overline{r}_T = \frac{1}{T}\sum_{t=1}^{T} r(I_t, J_t) \qquad \text{and} \qquad \overline{s}_T = \frac{1}{T}\sum_{t=1}^{T} s(I_t, J_t)$$

as $T \to \infty$.

Auxiliary quantities: the regrets. The respective regrets R_T and S_T of A and B till round T are defined as in (1.1) –up to the remplacement of the losses by payoffs– in terms of the best constant action of a given player all things being equal:

$$R_T = \max_{i\in\mathcal{A}} \sum_{t=1}^{T} r(i, J_t) - \sum_{t=1}^{T} r(I_t, J_t) \qquad \text{and} \qquad S_T = \max_{j\in\mathcal{B}} \sum_{t=1}^{T} s(I_t, j) - \sum_{t=1}^{T} s(I_t, J_t)\,.$$

The corresponding per-round regrets are denoted by

$$\overline{R}_T = \max_{i\in\mathcal{A}} r\big(i, \overline{\boldsymbol{q}}_T\big) - \overline{r}_T \qquad \text{and} \qquad \overline{S}_T = \max_{j\in\mathcal{B}} s\big(\overline{\boldsymbol{p}}_T, j\big) - \overline{s}_T\,. \tag{2.1}$$

By linearity, the maxima over $i \in \mathcal{A}$ and $j \in \mathcal{B}$ in the definitions above can be replaced by maxima over $\boldsymbol{p} \in \Delta(\mathcal{A})$ and $\boldsymbol{q} \in \Delta(\mathcal{B})$.

Minimize the regret in order to get a large per-round payoff? The interpretation of the notion of regret is less clear than in the setting of convex aggregation studied in the previous chapter, where some approximation error had to traded off with some estimation error (given by the regret). Therein, since the forecasts of the statistician do not influence the environment, one can define an intrinsic notion of best convex combination and could exploit oracle knowledge if some was available. But in the model of this chapter, if at each round the player A had played the optimal action against the realized sequence of actions $J_1, \dots, J_T$, the latter would have been different! We already underlined this issue in Section 1.1.3.

The defense of the notion of regret will rather rely on arguments of convergence in some sense towards sets of equilibria, where this convergence takes place as soon as both players ensure that their per-round regrets are small. Now, by definition, in a situation of equilibrium each player obtains his maximal mean payoff (where the maximality depends on the notion of equilibrium at hand).

In this respect the very interest of the notion of regret is that it is a quantity that each player can minimize on his own without the real need of interacting with his adversary; in particular no assumption on the rationality or the will to cooperate of the latter is required.

Outline of this introductory section. We first indicate how players can minimize their regrets and then study the consequences of a simultaneous minimization of the regrets: in general, convergence of the empirical frequencies of played action profiles towards the set of Hannan equilibria is achieved. In the special case of a zero-sum game one gets even stronger convergence results: the per-round payoffs tend to the value of the game and the pairs of empirical frequencies of played actions tend to the set of minimax equilibria (which corresponds here to the set of Nash equilibria).

2.1.1 Each player can control his regret, independently of the other player

We provide a straightforward adaptation of the results of the previous chapter to the case of payoffs instead of losses. We denote by $\|r\|_\infty$ an upper bound on the function $|r|$, we let

$$\eta_t = \frac{1}{\|r\|_\infty} \sqrt{\frac{8 \ln N}{t-1}}$$

for all $t \geqslant 2$, and we assume that player A, for instance, uses the following strategy: he chooses the action I_1 at random according to the uniform distribution $\boldsymbol{p}_1$ over $\mathcal{A}$, and for all $t \geqslant 2$, draws his actions I_t at random according to the probability distribution $\boldsymbol{p}_t$ whose components are given by

$$p_{i,t} = \frac{\exp\left(\eta_t \sum_{s=1}^{t-1} r(i, J_s)\right)}{\sum_{k \in \mathcal{A}} \exp\left(\eta_t \sum_{s=1}^{t-1} r(k, J_s)\right)} \tag{2.2}$$

for all $i \in \mathcal{A}$. An easy adaptation of the calculations around (1.22) then shows the following deterministic (i.e., with probability 1) inequality between random quantities: for all strategies τ of player B,

$$\sum_{t=1}^{T} r(\boldsymbol{p}_t, J_t) \geqslant \max_{i\in\mathcal{A}} \sum_{t=1}^{T} r(i, J_t) - \|r\|_\infty \sqrt{2T \ln N}\,.$$

In particular, the Hoeffding–Azuma inequality implies that at each round T and with probability at least $1-\delta$,

$$\overline{r}_T \geqslant \max_{i\in\mathcal{A}} r(i, \overline{\boldsymbol{q}}_T) - \|r\|_\infty \left(\sqrt{\frac{2}{T} \ln N} + \sqrt{\frac{1}{2T} \ln \frac{1}{\delta}} \right) . \tag{2.3}$$

Via a final application of the Borel–Cantelli lemma, we conclude that for all strategies τ of player B,

$$\liminf_{T\to\infty} \left\{ \overline{r}_T - \max_{i\in\mathcal{A}} r(i, \overline{\boldsymbol{q}}_T) \right\} \geqslant 0 \quad \text{a.s.}, \qquad \text{that is,} \qquad \limsup_{T\to\infty} \overline{R}_T \leqslant 0 \quad \text{a.s.} \tag{2.4}$$

All convergence results developed later in this chapter rely only on asymptotic statements of the form of (2.4); the statements like (2.3) give an idea of the rates of convergence. The strategies ensuring (2.4) are still called regret-minimizing strategies in this chapter.

Conclusions in terms of upper bounds on the regret. Player A has a strategy that solely relies on the knowledge of the payoff function r and on the observation of the actions J_t of his adversary player and that has a small regret as asserted by (2.3). In particular, player A does not have to know the other player's payoff function s nor to assume anything (like bounded rationality) on the strategy τ followed by player B.

The literature calls such strategies myopic: player A only pays attention to quantities that are close to him –namely, his own payoffs– and discards somehow farer away quantities –e.g., the strategy τ of his adversary. The rest of this chapter is devoted to illustrating the interest of such seemingly crude strategies: when the players minimize simultaneously their regrets, a convergence towards a set of equilibria takes place.

Other regret-minimizing strategies. We mention only two other families of strategies ensuring (2.4). Playing at each round the action with the best cumulative payoff so far can be disastrous as the associated regret can be linearly large. But a simple twist proposed by Hannan [Han57] leads to the minimization of the regret: the so-called "follow the perturbed leader" strategies add random perturbations to the cumulative payoffs and pick the action with the best perturbed cumulative payoff. They were recently re-introduced and re-studied by [KV03] and other researchers.

Another family proceeds from Blackwell's approachability theorem [Bla56], which we recall in the insert below. Here also the analysis was recently re-visited by [CBL03].

Insert: The approachability theorem**.**

Let $m : \mathcal{A}\times\mathcal{B} \to \mathbb{R}^d$ be a vector function, which is linearly extended on $\Delta(\mathcal{A})\times\Delta(\mathcal{B})$. Players A and B play repeatedly together, choosing simultaneously at each round $t \geqslant 1$ respective actions $I_t \in \mathcal{A}$ and $J_t \in \mathcal{B}$.

A set $\mathcal{C} \subset \mathbb{R}^d$ is said m–approachable by player A if the latter has a strategy σ such that, for all strategies τ of player B,

$$\lim_{T\to\infty} \inf_{c\in\mathcal{C}} \left\| c - \frac{1}{T}\sum_{t=1}^{T} m(I_t, J_t) \right\| = 0 \qquad \text{a.s.} \tag{2.5}$$

The following characterization of the approachability of closed convex sets follows from Neumann's minimax theorem (a special case of Sion's lemma, see Definition/Theorem 2.4). Even better, Blackwell also provided a strategy that approaches $\mathcal{C}$; it relies on convex projections and requires solving a linear program at each round.

Theorem 2.1 (Reference: [Bla56, Theorem 3]). *A closed convex set $\mathcal{C}$ of $\mathbb{R}^d$ is m–approchable if and only if*

$$\forall \boldsymbol{q} \in \Delta(\mathcal{B}), \quad \exists \boldsymbol{p} \in \Delta(\mathcal{A}), \qquad m(\boldsymbol{p}, \boldsymbol{q}) \in \mathcal{C}\,.$$

The existence of a regret-minimizing strategy then follows from the consideration of the non-positive orthant $\mathcal{C} =]-\infty, 0]^N$ and of the vector function m defined by

$$m(i,j) = \big(r(k,j) - r(i,j)\big)_{k\in\mathcal{A}}$$

for all $i \in \mathcal{A}$ and $j \in \mathcal{B}$.

2.1.2 Convergence towards the set of Hannan equilibria

The strategy of Hannan [Han57] is actually anterior to the one of Blackwell [Bla56] and it was the first one to ensure (2.4). This is why the following set of equilibria was named in honor of Hannan.

It is in terms of joint distributions $\pi = \big(\pi(i,j)\big)_{(i,j)\in\mathcal{A}\times\mathcal{B}}$; we denote by $\Delta(\mathcal{A}\times\mathcal{B})$ the simplex of all possible joint distributions.

Definition 2.2. The set of Hannan equilibria of a two-player game is given by the following (non empty) set of joint distributions:

$$\mathcal{H} = \Bigg\{ \pi \in \Delta(\mathcal{A}\times\mathcal{B}) : \quad \forall i \in \mathcal{A}, \ \sum_{k,\ell} \pi(k,\ell)\, r(k,\ell) \geqslant \sum_{k,\ell} \pi(k,\ell)\, r(i,\ell)$$
$$\text{and} \quad \forall j \in \mathcal{B}, \ \sum_{k,\ell} \pi(k,\ell)\, s(k,\ell) \geqslant \sum_{k,\ell} \pi(k,\ell)\, s(k,j) \Bigg\}.$$

Such a joint distribution π can indeed be interpreted as an equilibrium: suppose that an action profile (I, J) is drawn at random according to π by a mediator and that each player is recommended to play the action I or J that has been drawn for him. Then, in average, if a player abides by this recommendation, the other player has no incentive to replace his recommended action by another one that he would have set beforehand (i.e., before accessing the recommendation). Put differently, there are no profitable unilateral deviations given by fixed-in-advance actions.

An addition assumption on the game (its zero-sum character: $r + s = 0$) will be needed to ensure convergence results on the pairs $(\overline{\boldsymbol{p}}_T, \overline{\boldsymbol{q}}_T)$ of separate empirical distributions of actions taken. For the time being the quantities of interest are given by the empirical distributions of action profiles,

$$\overline{\pi}_T = \frac{1}{T} \sum_{t=1}^{T} \delta_{(I_t, J_t)} \, ;$$

their marginal distributions are $\overline{\boldsymbol{p}}_T$ and $\overline{\boldsymbol{q}}_T$. We assume that both players minimize their regrets, which rewrites as

$$\liminf_{T\to\infty} \left\{ \overline{r}_T - \max_{i\in\mathcal{A}} r\big(i, \overline{\boldsymbol{q}}_T\big) \right\} \geqslant 0 \quad \text{a.s.} \qquad \text{and} \qquad \liminf_{T\to\infty} \left\{ \overline{s}_T - \max_{j\in\mathcal{B}} s\big(\overline{\boldsymbol{p}}_T, j\big) \right\} \geqslant 0 \quad \text{a.s.}$$

The defining conditions of $\mathcal{H}$ being given by closed constraints, the above asymptotic inequalities ensure that each limit point π of the sequence of the $\overline{\pi}_T$ is a Hannan equilibrium: $\pi \in \mathcal{H}$. Now, since the set $\Delta(\mathcal{A} \times \mathcal{B})$ of all joint distributions is compact, a proof by contradiction finally shows that the sequence of the $\overline{\pi}_T$ indeed converges towards $\mathcal{H}$. We underline that the convergence takes places towards $\mathcal{H}$ and not necessarily towards a given point in $\mathcal{H}$ (the limit points do no necessarily have a unique value).

Proposition 2.3. When both players minimize their regrets, the sequence $(\overline{\pi}_T)$ of the empirical distributions of action profiles converges almost surely towards the set $\mathcal{H}$ of Hannan equilibria.

2.1.3 Zero-sum games: convergence towards the set of minimax equilibria

This section considers the special case of zero-sum games, where $r + s = 0$ or, put differently, $r = -s$. That is, the players have strictly opposite incentives. When both players know that they are playing such a game, they may use a simple optimal strategy which we describe next.

Some elementary results on zero-sum games

To state them we first need to define the value of a game; its definition follows from a special case of Sion's lemma, called von Neumann's minimax theorem.

Definition/Theorem 2.4 (von Neumann, 1928). The equalities

$$v \;=\; \max_{\boldsymbol{p}\in\Delta(\mathcal{A})}\ \min_{\boldsymbol{q}\in\Delta(\mathcal{B})}\ r(\boldsymbol{p},\boldsymbol{q}) = \min_{\boldsymbol{q}\in\Delta(\mathcal{B})}\ \max_{\boldsymbol{p}\in\Delta(\mathcal{A})}\ r(\boldsymbol{p},\boldsymbol{q})$$

define the value of a zero-sum game $r : \mathcal{A}\times\mathcal{B}\to\mathbb{R}$.

This notion leads by definition to the set of minimax equilibria (which are in fact the instantiation to this special case of the more general and historically posterior concept of Nash equilibria). In the next definition we identify the pairs $(\boldsymbol{p},\boldsymbol{q})\in\Delta(\mathcal{A})\times\Delta(\mathcal{B})$ of probability distributions with the product distributions that they induce in $\Delta(\mathcal{A}\times\mathcal{B})$.

Definition/Theorem 2.5. The set $\mathcal{N}$ of minimax equilibria of a zero-sum game $r : \mathcal{A}\times\mathcal{B}\to\mathbb{R}$ is given by the following (non empty) set of product distributions:

$$\begin{aligned}\mathcal{N} &= \mathcal{H}\cap(\Delta(\mathcal{A})\times\Delta(\mathcal{B}))\\ &= \Big\{(\boldsymbol{p},\boldsymbol{q}):\ \ \forall i\in\mathcal{A},\ \ r(\boldsymbol{p},\boldsymbol{q})\geqslant r(i,\boldsymbol{q})\quad\text{and}\quad \forall j\in\mathcal{B},\ \ r(\boldsymbol{p},\boldsymbol{q})\leqslant r(\boldsymbol{p},j)\Big\}.\end{aligned}$$

In particular, each pair $(\boldsymbol{p},\boldsymbol{q})\in\mathcal{N}$ achieves the value of the game: $r(\boldsymbol{p},\boldsymbol{q}) = v$.

Fix a pair of distributions $(\boldsymbol{p}^\star,\boldsymbol{q}^\star)$ in $\mathcal{N}$. If player B resorts to $\boldsymbol{q}_t = \boldsymbol{q}^\star$ at all rounds, that is, if he draws his actions J_t all independently at random according to the same distribution $\boldsymbol{q}^\star$, then the per-round payoff of all strategies σ of player A is bounded from above by v: indeed,

$$\limsup_{T\to\infty}\ \overline{r}_T = \limsup_{T\to\infty}\ \frac{1}{T}\sum_{t=1}^{T} r(I_t,J_t) = \limsup_{T\to\infty}\ r\big(\overline{\boldsymbol{p}}_T,\,\boldsymbol{q}^\star\big)\leqslant v \qquad \text{a.s.},$$

where we applied the Hoeffding–Azuma inequality as well as Borel–Cantelli lemma to get the second equality. Of course, the inequality in the display above can be an equality, for instance, when A draws his actions all independently at random according to the same distribution $\boldsymbol{p}^\star$. A similar result holds when A resorts to $\boldsymbol{p}_t = \boldsymbol{p}^\star$ at all rounds and when the strategies τ of B are under inspection: in this case player A obtains at least v as his asymptotic per-round payoff.

Conclusion and limitation. Whenever a player knows that the game is zero sum, he can use a strategy that is optimal in the sense that it guarantees an asymptotic per-round payoff that is as large as possible in the worst case; this strategy consists of computing a minimax equilibrium $(\boldsymbol{p}^\star,\boldsymbol{q}^\star)$ and drawing his actions in an independent and identically distributed fashion, according to his marginal distribution of this equilibrium.

The limitation is of course that the setting of interest is myopic –as indicated in the previous section– and that players only know their own payoff function; they in particular ignore whether the game is zero sum or not.

Convergence towards the set of minimax equilibria

We now show that the stated limitation can be circumvented: when both players use regret-minimizing strategies (which can be myopic and work independently) the results of the previous section hold. Namely, on the one hand player A (respectively, B) cannot have more than v as an asymptotic per-round payoff (respectively, more than $-v$) but on the other hand he can ensure that the latter is at least v (respectively, $-v$). These asymptotic per-round payoffs thus equal exactly v and $-v$.

Moreover we show that not only the sequence $(\overline{\pi}_T)$ converges towards $\mathcal{H}$ but also that the sequence $((\overline{\boldsymbol{p}}_T, \overline{\boldsymbol{q}}_T))$ of its marginal distributions converges towards $\mathcal{N}$.

Each player can guarantee the value of the game. When (2.4) is satisfied, the per-round payoffs of player A are in particular such that

$$\liminf_{T\to\infty} \overline{r}_T \geqslant \liminf_{T\to\infty} \max_{i\in\mathcal{A}} r\big(i, \overline{\boldsymbol{q}}_T\big) = \liminf_{T\to\infty} \max_{\boldsymbol{p}\in\Delta(\mathcal{A})} r\big(\boldsymbol{p}, \overline{\boldsymbol{q}}_T\big) \geqslant v \qquad \text{a.s.}\,; \tag{2.6}$$

by symmetry, when player B minimizes his regret, his per-round payoffs satisfy

$$\liminf_{T\to\infty} -\overline{r}_T \geqslant \liminf_{T\to\infty} \max_{\boldsymbol{q}\in\mathcal{B}} -r\big(\overline{\boldsymbol{p}}_T, \boldsymbol{q}\big) \geqslant -v \qquad \text{a.s.},$$

$$\text{that is,} \qquad \limsup_{T\to\infty} \overline{r}_T \leqslant \limsup_{T\to\infty} \min_{\boldsymbol{q}\in\Delta(\mathcal{B})} r\big(\overline{\boldsymbol{p}}_T, \boldsymbol{q}\big) \leqslant v \qquad \text{a.s.} \tag{2.7}$$

We therefore proved the following result.

Proposition 2.6. When both players of a zero-sum game with value v minimize their regrets, the almost-sure convergence $\overline{r}_T \to v$ takes place as $T \to \infty$.

Consequences. The above proposition actually shows that all inequalities in (2.6) and in (2.7) are equalities; in particular, a sandwich argument yields

$$\lim_{T\to\infty} r\big(\overline{\boldsymbol{p}}_T, \overline{\boldsymbol{q}}_T\big) = v \quad \text{a.s.}, \qquad \text{hence} \qquad \lim_{T\to\infty} \overline{r}_T - r\big(\overline{\boldsymbol{p}}_T, \overline{\boldsymbol{q}}_T\big) = 0 \quad \text{a.s.}$$

By getting back to the fact (2.4) that both players minimize their regrets, we get

$$\liminf_{T\to\infty} \left\{ r\big(\overline{\boldsymbol{p}}_T, \overline{\boldsymbol{q}}_T\big) - \max_{i\in\mathcal{A}} r\big(i, \overline{\boldsymbol{q}}_T\big) \right\} \geqslant 0 \quad \text{a.s.}$$

$$\text{and} \qquad \limsup_{T\to\infty} \left\{ r\big(\overline{\boldsymbol{p}}_T, \overline{\boldsymbol{q}}_T\big) - \min_{j\in\mathcal{B}} r\big(\overline{\boldsymbol{p}}_T, j\big) \right\} \leqslant 0 \quad \text{a.s.}$$

The same proof techniques as for Proposition 2.3 –namely, the fact that $\mathcal{N}$ is defined by closed constraints together with a compacity argument and a proof by contradiction– lead to the following statement of convergence, which is, here again, towards a set and not towards a given point of the set.

Proposition 2.7. In a zero-sum game, when both players minimize their regrets, the sequence $((\overline{\boldsymbol{p}}_T, \overline{\boldsymbol{q}}_T))$ of the pairs of empirical distributions of actions taken by each player converges almost surely towards the set $\mathcal{N}$ of minimax equilibria.

2.2 Regret minimization in games with partial monitoring [3, 6]

We deal in this section with the setting where one of the players –say player A– has only a partial monitoring of the actions taken by his opponent player. Formally, we introduce in addition to the notation and objects considered above a finite set $\mathcal{S}$ of signals and a feedback function $H : \mathcal{A} \times \mathcal{B} \to \Delta(\mathcal{S})$; the latter associates with each pair (i, j) of actions in $\mathcal{A} \times \mathcal{B}$ a probability distribution $H(i, j)$ over $\mathcal{S}$. The function H is linearly extended on $\Delta(\mathcal{A}) \times \Delta(\mathcal{B}) \to \Delta(\mathcal{S})$ as follows: for all distributions $\boldsymbol{p} \in \Delta(\mathcal{A})$ and $\boldsymbol{q} \in \Delta(\mathcal{B})$,

$$H(\boldsymbol{p}, \boldsymbol{q}) = \sum_{i \in \mathcal{A}} \sum_{j \in \mathcal{B}} p_i \, q_j \, H(i, j) \ \in \Delta(\mathcal{S}) \, .$$

Protocol of the repeated game with partial monitoring. We consider the viewpoint of player A. At each round $t \geqslant 1$, players A and B choose respective actions I_t and J_t, possibly at random according to probability distributions $\boldsymbol{p}_t$ and $\boldsymbol{q}_t$ that can depend on the information collected in the past. Player A obtains the payoff $r(I_t, J_t)$ but does not get to see neither J_t nor even the value of his payoff: he only observes a random variable K_t drawn independently at random according to $H(I_t, J_t)$. Player B –on the contrary– has a full monitoring: he observes I_t.

Outline of this section. We only extend the results of Section 2.1.3 to this setting of partial monitoring. More precisely, we introduce and defend an extension of the notion of regret such that whenever the corresponding criterion is minimized by all players and the game is zero sum, then the convergences stated in Propositions 2.6 and 2.7 still take place. The extended notion of regret of course also applies to the general case of finite games but for the sake of simplicity we focus on the special case of zero-sum games.

2.2.1 Extension of the notion of regret

Indistinguishability of certain randomized actions. Player A cannot distinguish between all probability distributions used by player B to draw his actions: two distributions $\boldsymbol{q}$ and $\boldsymbol{q}'$ with the equalities $H(i, \boldsymbol{q}) = H(i, \boldsymbol{q}')$ for all $i \in \mathcal{A}$ are identical to him as far as the received feedback is concerned. The short-hand notation for these equalities is $H(\,\cdot\,, \boldsymbol{q}) = H(\,\cdot\,, \boldsymbol{q}')$, where we defined the vector of probability distributions induced over the signals by a probability distribution $\boldsymbol{q}$ over $\mathcal{B}$ as

$$H(\,\cdot\,, \boldsymbol{q}) = \big(H(i, \boldsymbol{q})\big)_{i \in \mathcal{A}} \ \in (\Delta(\mathcal{S}))^{\mathcal{A}} \, .$$

We denote by

$$\mathcal{V} = \Big\{ H(\,\cdot\,, \boldsymbol{q}), \quad \boldsymbol{q} \in \Delta(\mathcal{B}) \Big\}$$

the set of these vectors as $\boldsymbol{q}$ varies. A generic element of $\mathcal{V}$ will usually be referred to as $\underline{h}$ in the rest of this section.

Because of the above-mentioned indistinguishability we introduce the mapping

$$\rho : \ (\boldsymbol{p}, \underline{h}) \in \Delta(\mathcal{A}) \times \mathcal{V} \longmapsto \min\Big\{ r(\boldsymbol{p}, \boldsymbol{q}) : \ \boldsymbol{q} \in \Delta(\mathcal{B}) \text{ such that } H(\,\cdot\,, \boldsymbol{q}) = \underline{h} \Big\} \in \mathbb{R}\,.$$

It indicates the minimal expected payoff when player A draws his action at random according to $\boldsymbol{p}$ and when player B does so with a probability distribution $\boldsymbol{q}$ inducing the vector of probability distributions over the signals $\underline{h}$. The function ρ is concave in its argument in $\Delta(\mathcal{A})$ and is convex in the one in $\mathcal{V}$.

Rewriting of the value of the game. The first key observation is as follows. In a zero-sum game, the value can be rewritten in terms of ρ as

$$v \ = \ \max_{\boldsymbol{p} \in \Delta(\mathcal{A})} \ \min_{\boldsymbol{q} \in \Delta(\mathcal{B})} \ r(\boldsymbol{p}, \boldsymbol{q}) = \ \max_{\boldsymbol{p} \in \Delta(\mathcal{A})} \ \min_{\underline{h} \in \mathcal{V}} \ \rho(\boldsymbol{p}, \underline{h}) = \ \min_{\underline{h} \in \mathcal{V}} \ \max_{\boldsymbol{p} \in \Delta(\mathcal{A})} \ \rho(\boldsymbol{p}, \underline{h})\,, \qquad (2.8)$$

where the first two equalities hold by definition of ρ while the third equality follows from a direct application of a generalized minimax theorem (that can still be expressed as a special case of Sion's lemma).

A suitable notion of regret. Given this rewriting of the value of the game we now aim at defining an extended notion of regret with which the guarantee given by (2.6) still holds: player A should be able to impose by minimizing his regret that his asymptotic per-round payoff is at least equal to v. Given the equalities above, to do so it suffices for player A to ensure that for all strategies τ of player B,

$$\liminf_{T \to \infty} \left\{ \overline{r}_T - \max_{\boldsymbol{p} \in \Delta(\mathcal{A})} \ \rho\Big(\boldsymbol{p},\, H\big(\,\cdot\,, \overline{\boldsymbol{q}}_T\big)\Big) \right\} \geqslant 0 \qquad \text{a.s.} \qquad (2.9)$$

We thus define the (per-round) regret of A on the first T rounds in the setting of partial monitoring as

$$\overline{R}_T^{\mathrm{PM}} = \max_{\boldsymbol{p} \in \Delta(\mathcal{A})} \ \rho\Big(\boldsymbol{p},\, H\big(\,\cdot\,, \overline{\boldsymbol{q}}_T\big)\Big) - \overline{r}_T\,.$$

In this section minimizing the regret and ensuring (2.9) mean that $\limsup_{T \to \infty} \overline{R}_T^{\mathrm{PM}} \leqslant 0$ a.s.

2.2.2 Results anterior to our contributions

The fundamental result stated below is an additional defense of the notion of regret just introduced: not only a player only needs to minimize this regret to guarantee the value of the game as his asymptotic per-round payoff but also it is indeed possible to do so. (It would not be the case here, in general, for the original notion of regret $\overline{R}_T$.)

Theorem 2.8 (Reference: [Rus99]). *There exists a strategy for player A with vanishing per-round regret:*

$$\limsup_{T \to \infty} \overline{R}_T^{\mathrm{PM}} \leqslant 0 \qquad \text{a.s.}$$

As states [Rus99] in the conclusion an open question was to exhibit an explicit regret-minimizing strategy –an expected by-product of this being to provide a simpler and constructive proof of Theorem 2.8. The original proof of the latter indeed relies on an abstract approachability theorem stated by [MSZ94] and with which (in contrast to Theorem 2.1) no natural strategy is associated. In addition the intuition was that the study of such an explicit strategy would lead to convergence-rates results (explicit upper bounds on the per-round regrets). A final desired property is that this explicit strategy admits a computationally efficient implementation. These various aims motivated the contributions posterior to [Rus99], which we now review.

Use of the simplest version of the approachability theorem in a special case. [MS03] designed a simple and explicit strategy relying on the approachability theorem (Theorem 2.1) in the case where the feedback function H does not depend on the actions of player A; but in the general case this strategy ensures a weaker result than the one of (2.9).

Detailed overview of the case of sufficient feedback

We consider in this section the games with partial monitoring such that the aims (2.4) and (2.9) coincide, that is, such that for all probability distributions $\boldsymbol{q}$ over $\mathcal{B}$,

$$\max_{\boldsymbol{p}\in\Delta(\mathcal{A})} \rho(\boldsymbol{p},\, H(\,\cdot\,,\boldsymbol{q})) = \max_{\boldsymbol{p}\in\Delta(\mathcal{A})} r(\boldsymbol{p},\boldsymbol{q}) = \max_{i\in\mathcal{A}} r(i,\boldsymbol{q})\,.$$

Notion of sufficient feedback. The word "sufficient" here has the same meaning as, e.g., in the notion of sufficient statistics. In the present case it in particular holds true that the equality of the distributions over the signals $H(\,\cdot\,,\boldsymbol{q}) = H(\,\cdot\,,\boldsymbol{q}')$ entails the equalities of the target quantities

$$\max_{i\in\mathcal{A}} r(i,\boldsymbol{q}) = \max_{i\in\mathcal{A}} r(i,\boldsymbol{q}')\,.$$

Examples of such games are, for instance, the games in which H reveals the probability distribution chosen by player B, that is, feedback functions satisfying the following property: $H(\,\cdot\,,\boldsymbol{q}) = H(\,\cdot\,,\boldsymbol{q}')$ if and only if $\boldsymbol{q} = \boldsymbol{q}'$. This property can be equivalently characterized in terms of a certain matrix representation of H being of full rank.

Reconstruction of the payoff function in terms of the feedback function. A more general situation arises when the payoff function r can be reconstructed from the feedback function H. It was made formal by [PS01]; we only discuss it in the simpler case where for all pairs $(i,j) \in \mathcal{A}\times\mathcal{B}$, the probability distribution $H(i,j)$ over the signals is a Dirac mass, on a signal that we denote by $h(i,j)$. With no loss of generality we then re-encode the set of signals $\mathcal{S}$ into a finite subset of $[0,1]$; the reconstruction condition can then stated as the existence of a function $f : \mathcal{A}\times\mathcal{A} \to \mathbb{R}$ such that

$$\forall\,(i,j)\in\mathcal{A}\times\mathcal{B}, \qquad r(i,j) = \sum_{k\in\mathcal{A}} f(i,k)\,h(k,j)\,. \tag{2.10}$$

[PS01] actually shows that all games with sufficient feedback can be algorithmically cast into to the above-mentioned framework of deterministic feedback associated with the reconstruction property (2.10); this reduction is performed via suitable elementary transformations of the underlying game (like the duplication of actions and/or signals).

Estimation of the unobserved payoffs. Player A may then estimate at each round his own payoff $r(I_t, J_t)$ and the payoffs $r(i, J_t)$ that he would have obtained by playing other actions $i \in \mathcal{A}$, based on the sole piece of information he receives: the deterministic feedback $K_t = h(I_t, J_t)$. More precisely, for all actions $i \in \mathcal{A}$, he computes the statistic

$$\widehat{r}_{i,t} = \frac{f(i, I_t)\, K_t}{p_{I_t,t}} = \frac{f(i, I_t)\, h(I_t, J_t)}{p_{I_t,t}} \,,$$

where the action I_t was drawn at random according to the probability distribution $\boldsymbol{p}_t$ (which we assume had full support $\mathcal{A}$) and where $p_{I_t,t}$ denotes the I_t–th component of $\boldsymbol{p}_t$. These statistics are conditionally unbiased:

$$\mathbb{E}\Big[\, \widehat{r}_{i,t} \,\Big|\, \boldsymbol{p}_t, J_t \Big] = \sum_{k \in \mathcal{A}} \frac{f(i,k)\, h(k, J_t)}{p_{k,t}}\, p_{k,t} = r(i, J_t) \,,$$

where we used the reconstruction property (2.10).

[ACBFS02] was actually the first to propose such an unbiased estimation of unobserved payoffs, in the simpler case of multi-armed bandit problems of Chapter 4. (We also detailed posterior similar estimation techniques for unobserved losses in Section 1.3.1.)

Associated strategy. The strategy then proposed by [PS01] generalizes the exponentially weighted average strategy (2.2) of the case of full monitoring; it resorts, for all rounds $t \geqslant 2$, to probability distributions $\boldsymbol{p}_t$ whose components are given by

$$p_{i,t} = (1 - \gamma_t) \frac{\exp\Big(\eta_t \sum_{s=1}^{t-1} \widehat{r}_{i,s}\Big)}{\sum_{k \in \mathcal{A}} \exp\Big(\eta_t \sum_{s=1}^{t-1} \widehat{r}_{k,s}\Big)} + \frac{\gamma_t}{N} \tag{2.11}$$

for all $i \in \mathcal{A}$, where $\eta_t > 0$ and $\gamma_t > 0$ are two parameters to be set by the analysis. The interpretation is, on the one hand, that the unobserved payoffs are replaced by their estimators and, on the other hand, that some minimal exploration of all actions is enforced via the mixing with the uniform distribution; in contrast, the first term in the right-hand side of the definition (2.11) is called the exploration term and a trade-off (measured by the value of γ_t) has to be made here between exploration and exploitation.

From a more technical viewpoint the lower bound of γ_t/N imposed by (2.11) is useful to control the deviations of the estimators $\widehat{r}_{i,t}$ around their conditional expectations. Based on this [PS01] proposes an upper bound on the original notion of regret $\overline{R}_T$ –the one defined in (2.1)– of the order of $T^{-1/4}$ up to logarithmic factors. (This is of course stronger than simply bounding the regret $\overline{R}_T^{\mathrm{PM}}$.)

2.2.3 Explicit and efficient general strategy [3, 6]

Our contributions took place in two steps.

First step: warm-up [3]. We first re-visited the results of [PS01] and showed that the analysis of the strategy described above could be improved to yield an upper bound of the order of $T^{-1/3}$ on the per-round regret $\overline{R}_T$. We also exhibited an example of partial information game with sufficient feedback in which this convergence rate towards 0 could not be improved –hence proving the optimality of the procedure as far as $\overline{R}_T$ was concerned. However, the constructive general minimization of $\overline{R}_T^{\mathrm{PM}}$ (when it is not possible to minimize $\overline{R}_T$) was left unsolved.

Second step: analysis in the general case [6] and minimization of $\overline{R}_T^{\mathrm{PM}}$. We realized after a while (and this was the key observation) that we should not focus too much on the individual payoffs $r(i, J_t)$ but only keep in mind the target quantity

$$\max_{\boldsymbol{p}\in\Delta(\mathcal{A})} \rho\Big(\boldsymbol{p},\, H\big(\,\cdot\,, \overline{\boldsymbol{q}}_T\big)\Big)\,.$$

It thus suffices to estimate $H\big(\,\cdot\,, \overline{\boldsymbol{q}}_T\big)$ and to do so we used here again the conditionally-unbiased estimation trick proposed by [ACBFS02] in the context of multi-armed bandit problems. Whenever it is necessary we identify the probability distributions over $\mathcal{S}$ with vectors in $\mathbb{R}^{\mathcal{S}}$.

Our estimator for the distribution $H(i, J_t)$ over the signals is the statistic

$$\widehat{h}_{i,t} = \frac{\delta_{K_t}}{p_{i,t}} \mathbb{I}_{\{I_t=i\}}\,,$$

where we recall that K_t denotes the feedback available at round t: a signal drawn at random according to $H(I_t, J_t)$. This estimator is conditionally unbiased with respect to the random variables $\boldsymbol{p}_t$ and J_t :

$$\begin{aligned}\mathbb{E}\Big[\widehat{h}_{i,t}\,\Big|\,\boldsymbol{p}_t, J_t\Big] = \frac{1}{p_{i,t}}\mathbb{E}\Big[\,\delta_{K_t}\mathbb{I}_{\{I_t=i\}}\,\Big|\,\boldsymbol{p}_t, J_t\Big] &= \frac{1}{p_{i,t}}\mathbb{E}\Big[\,H(I_t, J_t)\,\mathbb{I}_{\{I_t=i\}}\,\Big|\,\boldsymbol{p}_t, J_t\Big] \\ &= \frac{1}{p_{i,t}}\; p_{i,t}\, H(i, J_t) = H(i, J_t)\,,\end{aligned}$$

where we first considered expectations with respect to K_t and then with respect to I_t.

We now use the (Euclidian) convex projection operator onto $\mathcal{V}$, which we denote by Π. A concentration-of-the-measure argument applicable in Hilbert spaces [CW96] then shows that for a given large enough integer m and for all integers $b \geqslant 0$,

$$\underline{\widehat{h}}^b = \Pi\left(\frac{1}{m}\sum_{t=bm+1}^{(b+1)m}\Big[\widehat{h}_{i,t}\Big]_{i\in\mathcal{A}}\right) \quad \text{is a good estimator of} \quad \underline{h}^b = \frac{1}{m}\sum_{t=bm+1}^{(b+1)m} H(\,\cdot\,, J_t)\,. \tag{2.12}$$

Parameters: an integer $m \geqslant 1$, two real numbers η, $\gamma > 0$

Initialization: $\boldsymbol{w}^0 = (1, \dots, 1)$

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

1. Compute the integer b such that $bm + 1 \leqslant t \leqslant (b+1)m$;
2. Resort to the probability distribution $\boldsymbol{p}^b = (1-\gamma)\widetilde{\boldsymbol{p}}^b + \gamma \boldsymbol{u}$, where $\widetilde{\boldsymbol{p}}^b$ is defined component-wise by
$$\widetilde{p}_i^b = \frac{w_i^b}{\sum_{k \in \mathcal{A}} w_k^b}\,, \qquad \text{for } i \in \mathcal{A},$$
and where $\boldsymbol{u}$ denotes the uniform distribution over $\mathcal{A}$;
3. Draw the action I_t at random according to $\boldsymbol{p}_t = \boldsymbol{p}^b$;
4. Get the feedback K_t;
5. If $t = (b+1)m$, perform the update
$$w_i^{b+1} = w_i^b \exp\left(\eta \left(\partial\rho\Big(\boldsymbol{p}^b, \widehat{\underline{h}}^b\Big)\right)_i\right) \qquad \text{for } i \in \mathcal{A},$$
where $\widehat{\underline{h}}^b$ is defined in (2.12) and where $\partial\rho$ denotes a subgradient of ρ with respect to its first argument.

Figure 2.1. A strategy minimizing the regret in the case of a game with partial monitoring.

The remaining two ingredients to design our strategy are the following ones. First, a trade-off between exploitation and uniform exploration needs to set here as well. Second, the fact that the pessimistic payoff function ρ is concave and uniformly Lipschitz in its argument in $\Delta(\mathcal{A})$ entails a (uniform) linear upper bound of the form of (1.10). To that end, for all interior points $\boldsymbol{p}$ in $\Delta(\mathcal{A})$ and all elements $\underline{h}$ in $\mathcal{V}$, we denote by $\partial\rho(\boldsymbol{p}, \underline{h})$ a subgradient of $\rho(\,\cdot\,, \underline{h})$ at $\boldsymbol{p}$. This subgradient is a vector in $\mathbb{R}^N$ and we refer to its i–th component by a subscript i.

Our strategy is based on these three ingredients and is formally defined in Figure 2.1. It is simple and computationally efficient; it also provides a constructive proof of Theorem 2.8, together with convergence rates towards 0 of the per-round regret. We now detail somewhat informally these rates results; in [6] we of course indicated how to tune the parameters of the strategy and only stated explicit finite-time performance bounds, which we do not here for the sake of simplicity.

Theorem 2.9. *The strategy of Figure 2.1 –when tuned with adequate parameters– ensures that with probability at least* $1 - \delta$*,*

$$\overline{R}_T^{\mathrm{PM}} \leqslant \mathcal{O}\Big(T^{-1/5}\sqrt{\ln(T/\delta)}\Big)\,.$$

Simplified versions of this strategy entail in addition the following improvements on the convergence rates towards 0. These rates are indeed upper bounded with probability at least $1-\delta$ and up to a $\sqrt{\ln(T/\delta)}$ factor: by $T^{-1/4}$ in the case where the feedback is random but only depends on the action taken by player B; by $T^{-1/3}$ in the case where the feedback is deterministic but depends on the action profile chosen by players A and B; by $T^{-1/2}$ in the case where the feedback is deterministic and only depends on the action taken by player B. The minimax lower bounds stated in [CBFH$^+$97] and [3] also indicate that the convergence rates obtained in the above cases of deterministic feedback are optimal, up to logarithmic factors.

2.2.4 Results posterior to our contributions and perspectives for future research

Optimal convergence rates (in all cases) for other efficient strategies

In [6] we were unable to tell whether the convergence rates $T^{-1/5}$ and $T^{-1/4}$ exhibited in the cases of random feedback mentioned above were optimal or not. [Per09c] showed recently that this was not the case by constructing a strategy with regret bounded with probability at least $1-\delta$ by something of the order of $T^{-1/3}\ln(1/\delta)$ in all games of partial monitoring. A simplified variant of this strategy in the case of a random feedback only depending on the action taken by player B leads to the bound $T^{-1/2}\ln(1/\delta)$. These convergence rates are optimal as already discussed above. Moreover, the proposed strategies are also computationally efficient as they rely on a finite subset of the simplex $\Delta(\mathcal{A})$ containing for each vector $\underline{h} \in \mathcal{V}$ a best reply of player A to $\underline{h}$ in the sense of ρ; [Per09c] proves that such a finite subset always exists and provides insights to compute it algorithmically.

Extension of the approachability theorem

We described in the previous section how the approachability theorem (Theorem 2.1) guarantees the existence of regret-minimizing strategies in the case of a full monitoring. We discussed in this section the existence of regret-minimizing strategies in the case of partial monitoring; but this result corresponds to the approachability of some convex set for some vector payoff function. More precisely, as indicated by [Rus99], achieving (2.9) is ensuring that the closed convex set

$$\mathcal{C} = \left\{ (z, \boldsymbol{q}) \in \mathbb{R} \times \Delta(\mathcal{B}) : \quad z \geqslant \max_{\boldsymbol{p} \in \Delta(\mathcal{A})} \rho\big(\boldsymbol{p},\, H(\,\cdot\,, \boldsymbol{q})\big) \right\}$$

is approachable for the vector payoff function $m : \mathcal{A} \times \mathcal{B} \to \mathbb{R} \times \Delta(\mathcal{B})$ defined as follows: for all action profiles $(i, j) \in \mathcal{A} \times \mathcal{B}$,

$$m(i,j) = \left[\begin{array}{c} r(i,j) \\ \delta_j \end{array} \right].$$

[Per09a] states and proves an extension of the approachability theorem to games with partial monitoring –that however suffers from two drawbacks. First, the obtained

characterization of approachability comes without an efficient associated strategy (a direct implementation of the semi-explicit strategy proposed in his proof would require a computation time exponential in T, more information is provided in Section 2.4.3). Second, no rate is worked out for the convergence (2.5). It would be interesting to deal with these two issues.

Theorem 2.10 (Reference: [Per09a]). *Let $\mathcal{C} \subset \mathbb{R}^d$ be a closed convex set and $m : \mathcal{A} \times \mathcal{B} \to \mathbb{R}^d$ be a vector function. Then $\mathcal{C}$ is m–approachable if and only if*

$$\forall \underline{h} \in \mathcal{V}, \quad \exists \boldsymbol{p} \in \Delta(\mathcal{A}), \quad \forall \boldsymbol{q} \in \Delta(\mathcal{B}) \text{ such that } H(\,\cdot\,, \boldsymbol{q}) = \underline{h}, \qquad m(\boldsymbol{p}, \boldsymbol{q}) \in \mathcal{C}\,.$$

The proof of this deep result relies on the one hand on various concepts and technical elements developed in [6] and on the other hand on the existence of strategies minimizing the so-called internal regret in games with partial monitoring, where the mentioned existence follows from the existence of another class of strategies called calibrated strategies. The next sections of this chapter will deal with these two notions of calibration and internal regret.

> Remark in passing. Historically the first calibrated strategies were constructed based on strategies minimizing the internal regret of some auxiliary game with full monitoring and these calibrated strategies were the keystone to design strategies minimizing the internal regret in games with partial monitoring. Presenting the various strategies in the order in which they should be constructed the ones based on the other ones thus required switching between internal regret and calibration. But one of our recent contributions exhibits an intrinsic proof of calibration by approachability. This is why one can since then discuss first calibration and then the minimization of internal regret –in two distinct sections with no cross-reference–, which we do next.

2.3 Direct construction of calibrated strategies based on approachability [9]

In the game of calibration player A has to predict the actions of player B. The latter still picks his actions in a finite set denoted by $\mathcal{B}$ but the actions of player A are now given by the set $\Delta(\mathcal{B})$ of the probability distributions over $\mathcal{B}$. We equip $\Delta(\mathcal{B})$ with the topology induced by the canonical inclusion in $\mathbb{R}^{\mathcal{B}}$ and in particular consider the Borel σ–algebra generated by this topology.

Protocol of the repeated game. At each round players A and B choose simultaneously and based on past information respective actions $P_t \in \Delta(\mathcal{B})$ and $J_t \in \mathcal{B}$. These actions are actually drawn at random according to probability distributions ν_t over $\Delta(\mathcal{B})$ and $\boldsymbol{q}_t$ over $\mathcal{B}$.

Calibration aim. We fix a norm $\|\cdot\|$ on $\Delta(\mathcal{B})$, for instance, the ℓ^1–norm. The aim of player A is to design a strategy σ which delivers calibrated forecasts, that is, which

ensures that for all strategies of player B,

$$\forall \varepsilon > 0, \quad \forall \boldsymbol{p} \in \Delta(\mathcal{B}), \qquad \lim_{T \to +\infty} \left\| \frac{1}{T} \sum_{t=1}^{T} \mathbb{I}_{\left\{ \|P_t - \boldsymbol{p}\| \leqslant \varepsilon \right\}} \left(P_t - \delta_{J_t} \right) \right\| = 0 \qquad \text{a.s.} \qquad (2.13)$$

The quantity tending to 0 above is called the calibration error (at round T). The interpretation is as follows: player A wants to ensure that for all distributions $\boldsymbol{p}$ on the behavior of player B, the empirical distribution of the actions of the latter on rounds when the former had predicted a behavior close to $\boldsymbol{p}$ is indeed close to $\boldsymbol{p}$. It is a matter of an in-hindsight coherence of the forecasts of the distributions with respect to their realizations.

Calibrated strategies will be used in the next section as auxiliary strategies. In particular when player A actually has an own set of actions $\mathcal{A}$ and an own payoff function $r : \mathcal{A} \times \mathcal{B} \to \mathbb{R}$, he can choose his action $I_t \in \mathcal{A}$ at round t as a function of the forecast P_t of the behavior of player B output by an auxiliary calibrated strategy. Since its forecasts P_t are accurate in the sense of (2.13), the average payoff of player A is likely to exhibit interesting properties with this two-step prediction scheme.

Literature review. With his sense of humor Foster [Fos99] was already writing:

> "Over the past few years many proofs of the existence of calibration have been discovered. Each of the following provides a different algorithm and proof of convergence: Foster and Vohra [FV91, FV98]; Hart [Har95]; Fudenberg and Levine [FL99]; Hart and Mas-Colell [HMC00]. Does the literature really need one more? Probably not."

But despite all he then could publish his calibrated strategy for the binary case (i.e., where $\mathcal{B}$ contains only two elements) as it was more direct and shorter than all previously exhibited calibrated strategies which he listed exhaustively above. His strategy relies on the approachability theorem (Theorem 2.1). Actually all known calibrated strategies rely to some extent on approachability results, sometimes in an indirect or hidden manner or through auxiliary (sub)strategies. For instance, the calibrated strategy of [FV98] is based on an auxiliary internal-regret-minimizing strategy and the latter can be obtained in a natural way by approachability.

[FL99] and [HMC00] consider the case of action sets $\mathcal{B}$ with a finite (but arbitrary) number of outcomes; they do not exhibit convergence rates towards 0 of the calibration error (2.13). On the contrary, the strategies of [FV91, FV98, Fos99] are only valid for the binary case but lead to such convergence rates, of the order of $T^{-1/4}$ up to logarithmic factors.

Our contribution. We deal with the general case of a finite number of actions in $\mathcal{B}$ and exhibit a simple strategy –to our knowledge, the simplest strategy among all known calibrated strategies– based on a direct application of the approachability theorem. In this respect it captures the essence of the previous proofs of existence of calibrated

strategies. In addition we are able to work out explicit convergence rates towards 0 of the calibration error; as expected these rates depend on the cardinality of $\mathcal{B}$.

2.3.1 Preliminary construction of an ε–calibrated strategy

Most of the contributions mentioned above, e.g., [FV91, FV98, Fos99, FL99], do not tackle (2.13) directly and consider first a relaxed criterion called ε–calibration, where $\varepsilon > 0$ is a parameter chosen by player A. To that end they consider an ε–grid of $\Delta(\mathcal{B})$, that is, a finite set of distributions $\mathcal{G}_\varepsilon = \{\boldsymbol{p}_1, \dots, \boldsymbol{p}_{N_\varepsilon}\}$ such that the balls with centers $\boldsymbol{p}_k$ and radius ε cover $\Delta(\mathcal{B})$ as k varies in $\{1, \dots, N_\varepsilon\}$.

Definition 2.11. A strategy of player A is ε–calibrated if it only picks forecasts in some ε–grid $\mathcal{G}_\varepsilon$ and if it ensures that for all strategies τ of player B,

$$\limsup_{T\to+\infty} \sum_{k=1}^{N_\varepsilon} \left\| \frac{1}{T} \sum_{t=1}^{T} \mathbb{I}_{\{P_t = \boldsymbol{p}_k\}} (\boldsymbol{p}_k - \delta_{J_t}) \right\| \leqslant \varepsilon \qquad \text{a.s.} \tag{2.14}$$

We could then prove the following fundamental result.

Theorem 2.12. *With each ε–grid $\mathcal{G}_\varepsilon$ of $\Delta(\mathcal{B})$ can be associated an ε–calibrated strategy based on the approachability theorem.*

Proof. The game of interest is a finite game: the actions of player A are indexed by the finite set $\mathcal{G}_\varepsilon$ while the ones of player B are still given by the action set $\mathcal{B}$. We define the following vector function $m : \mathcal{G}_\varepsilon \times \mathcal{B} \to \mathbb{R}^{\mathcal{G}_\varepsilon \times \mathcal{B}}$, where we identify probability distributions over $\mathcal{B}$ with vectors in $\mathbb{R}^{\mathcal{B}}$: for all $k \in \{1, \dots, N_\varepsilon\}$ and $j \in \mathcal{B}$,

$$m(\boldsymbol{p}_k, j) = (\underline{0}, \dots, \boldsymbol{p}_k - \delta_j, \underline{0}, \dots, \underline{0}) ,$$

which is a vector composed by $k-1$ occurrences of the zero element $\underline{0} \in \mathbb{R}^{\mathcal{B}}$, followed by a non-null element in $\mathbb{R}^{\mathcal{B}}$, and completed by another series of $N_\varepsilon - k$ zero elements.

We now define the closed convex target set $\mathcal{C}$ as the closed ball centered at $(\underline{0}, \dots, \underline{0})$ and with radius ε for the norm $\|\cdot\|$. Now, the condition (2.14) of ε–calibration can be rewritten exactly as the fact that

$$\frac{1}{T} \sum_{t=1}^{T} m(P_t, J_t) = \left(\frac{1}{T} \sum_{t=1}^{T} \mathbb{I}_{\{P_t = \boldsymbol{p}_1\}} (\boldsymbol{p}_1 - \delta_{J_t}),\ \dots,\ \frac{1}{T} \sum_{t=1}^{T} \mathbb{I}_{\{P_t = \boldsymbol{p}_{N_\varepsilon}\}} (\boldsymbol{p}_{N_\varepsilon} - \delta_{J_t}) \right)$$

converges to $\mathcal{C}$ almost surely.

The existence of an ε–calibrated strategy is then equivalent to the m–approachability of $\mathcal{C}$, which we prove next by resorting to the characterization stated in Theorem 2.1. Let $\boldsymbol{q} \in \Delta(\mathcal{B})$ be a distribution over the actions of player B. By the defining properties of the ε–grid $\mathcal{G}_\varepsilon$, there exists $k \in \{1, \dots, N_\varepsilon\}$ such that $\|\boldsymbol{p}_k - \boldsymbol{q}\| \leqslant \varepsilon$, which in turn entails that

$$m(\boldsymbol{p}_k, \boldsymbol{q}) \in C .$$

(The distribution over $\mathcal{G}_\varepsilon$ of the approachability theorem can thus be taken as a Dirac mass.) □

Discussion of the memory and computational complexities. We state in [9] the strategy associated with Theorem 2.12: we first indicate how to compute at each round t the convex projection prescribed by the strategy canonically associated with the approachability theorem and second explain how the (approximate) solution of some linear program enables to associate with this projection a suitable probability distribution ν_t over $\mathcal{G}_\varepsilon$. Up to logarithmic factors the per-round complexity of the proposed implementation is of the the order of $\varepsilon^{-|\mathcal{B}|-1}$, where $|\mathcal{B}|$ denotes the cardinality of $\mathcal{B}$.

2.3.2 Construction of a calibrated strategy

By following the methodology sketched in [CBL06, Section 4.5 and Exercise 7.23], which relies on concentration-of-the-measure techniques in Hilbert spaces [CW96], we could prove the following result. To the best of our knowledge, it is the first convergence rates result for the calibration error when the action set $\mathcal{B}$ contains more than two elements. (In the statement of the theorem by playing a given strategy in blocks we mean using the doubling trick of Section 1.4.1 with this strategy.)

Theorem 2.13. *A strategy playing the strategies of Theorem 2.12 in blocks ensures that*

$$\limsup_{T\to\infty} \frac{T^{1/(|\mathcal{B}|+1)}}{\sqrt{\ln T}} \sup_{\boldsymbol{p}\in\Delta(\mathcal{B})} \sup_{\varepsilon>0} \left\| \frac{1}{T}\sum_{t=1}^{T} \mathbb{I}_{\left\{\|P_t-\boldsymbol{p}\|\leqslant\varepsilon\right\}} \left(P_t - \delta_{J_t}\right) \right\| \leqslant \Gamma_{|\mathcal{B}|} \qquad \text{a.s.},$$

where $\Gamma_{|\mathcal{B}|}$ denotes a constant that only depends on $|\mathcal{B}|$.

The obtained rates are uniform over the elements of $\Delta(\mathcal{B})$ and the ε–balls centered at them. Actually, a stronger uniformity holds: the calibration error can be defined in terms of indicator functions $P_t \in \mathcal{L}$, where $\mathcal{L}$ is a given Borel set, and the supremum in the definition of the uniform calibration error can then hold over all Borel sets.

2.3.3 Comparison with anterior and posterior contributions; perspectives for future research

Detailed comparisons with anterior and posterior contributions

In terms of convergence rates. The only anterior convergence rates result that is stated explicitly and that we are aware of is the following one. In the case where $|\mathcal{B}| = 2$ (only this case is worked out) [CBL06, Section 4.5] indicates how to obtain a (uniform) convergence rate for the calibration error of a strategy based on the ε–calibrated strategies exhibited by [FV98]. This rate is of the order of $T^{-1/4}$ up to logarithmic factors and it should be compared to the corresponding rate $T^{-1/3}$ provided by Theorem 2.13.

It did not seem easy to us to extend the strategy based on [FV98] to the non-binary case where $|\mathcal{B}| > 2$; but [Per10] did so, based on a significant modification of the base ε–calibrated strategies and by exploiting techniques developed in [9]. (A brief word on that, for experts only: the modified strategies do not minimize anymore their internal regrets with respect to all elements of the ε–grid, they do so only with respect to some nearest

neighbors.) He obtained the same convergence rates as in Theorem 2.13. He proved in addition that a rate of $T^{-1/2}$ –independent of the cardinality of $\mathcal{B}$– could be obtained for a uniform calibration error defined only in terms of a countable neighborhood base of $\Delta(\mathcal{B})$.

In terms of the memory and computational complexities of the ε–calibrated strategies. [Per10] does not study the complexity of the implementation of his strategy, which we recall is suited for the general case of a finite action set $\mathcal{B}$. What we can say is that this implementation requires memory and computational complexities which on the one hand are at least of the order of $\varepsilon^{1-|\mathcal{B}|}$ and on the other hand are bounded by something of the order of $\varepsilon^{2(1-|\mathcal{B}|)}$.

In the binary case, which is the most studied one, the best memory and computational complexities for an ε–calibrated strategy are of the order of $1/\varepsilon$ and are achieved by the simple and explicit strategy introduced by [Fos99]. This is to be compared to the complexities $1/\varepsilon^2$ and $1/\varepsilon^3$ obtained respectively by the strategies of [FV98] and [9] in this case.

Perspectives for future research

To the best of our knowledge no minimax lower bound result is available for the calibration error; such a lower bound could be either on the convergence rate of the calibration error towards 0 or on the (memory and/or computational) complexities required to implement ε–calibrated strategies. The latter may be obtained via a trade-off between the memory complexity and the computation complexity. The underlying open question is to determine whether there exist efficient calibrated strategies, i.e., whose complexities do not increase exponentially fast with the cardinality of $\mathcal{B}$ as is currently the case. The aim in terms of convergence rates is maybe clearer: one could hope that the convergence rates $T^{-1/(|\mathcal{B}|+1)}$ exhibited both by [9] and [Per10] are optimal up to logarithmic factors.

2.4 Convergence towards the set of correlated equilibria [1, 4]

After this (technical but useful) digression on calibration we get back to our main point: convergences towards sets of equilibria like the ones exhibited in Sections 2.1.2 and 2.1.3 and which justified the associated notions of regret. We consider again their associated framework and notation, at least to start this section.

2.4.1 Games with finite action sets $\mathcal{A}$ and $\mathcal{B}$

Another (and more popular) notion of equilibrium. The equilibrium notion of Section 2.1.2 is not much considered in game theory; preferred equilibrium notions are Nash equilibria – which the minimax equilibria are a special case of– and correlated equilibria. However the first equilibria are (NP–)hard to compute in general: therefore, one cannot expect

that there exist simple and efficient strategies such that, when all players follow them, a convergence (in some sense) to the set of Nash equilibria of the game takes place.

This is in contrast with correlated equilibria [Aum74, Aum87], which are, similarly to the Hannan equilibria, a subset of the joint distributions given by polynomially many (in the numbers of actions) linear constraints. These constraints are expressed in terms of functions $\varphi : \mathcal{A} \to \mathcal{A}$ and $\psi : \mathcal{B} \to \mathcal{B}$ called departure functions.

Definition 2.14. The set of correlated equilibria of a finite two-player game is given by the following (non empty) set of joint distributions:

$$\mathcal{E} = \left\{ \pi \in \Delta(\mathcal{A} \times \mathcal{B}) : \quad \forall \varphi : \mathcal{A} \to \mathcal{A}, \ \sum_{i,j} \pi(i,j)\, r(i,j) \geqslant \sum_{i,j} \pi(i,j)\, r\big(\varphi(i), j\big) \right.$$
$$\left. \text{and} \quad \forall \psi : \mathcal{B} \to \mathcal{B}, \ \sum_{i,j} \pi(i,j)\, s(i,j) \geqslant \sum_{i,j} \pi(i,j)\, s\big(i, \psi(j)\big) \right\}.$$

Such a joint distribution π can indeed be interpreted as an equilibrium: suppose that an action profile (I, J) is drawn at random according to π by a mediator and that each player is recommended to play the action I or J that has been drawn for him. Then, in average, if a player abides by this recommendation, the other player has no incentive to replace his recommended action by another one that he would choose only based on this recommendation (via a departure function). Put differently, there are no profitable unilateral deviations given by functions of the recommended actions.

This interpretation should be compared to the one provided after Definition 2.2: the only difference is that for correlated equilibria the deviation from the recommended action can be expressed as a function of the latter while for Hannan equilibria it has to be fixed beforehand, which corresponds to constant departure functions only. Therefore the inclusion $\mathcal{E} \subseteq \mathcal{H}$ holds, which shows that the aim stated below is in general more ambitious than the one pursued in Section 2.1.2 –but less ambitious, in the case of zero-sum games, than the aim of Section 2.1.3 since $\mathcal{N} \subseteq \mathcal{E}$.

Aim. The quantities of interest are the empirical distributions of the action profiles,

$$\overline{\pi}_T = \frac{1}{T} \sum_{t=1}^{T} \delta_{(I_t, J_t)}\,,$$

and we will show that when both players minimize simultaneously their so-called internal regrets, the sequence $(\overline{\pi}_T)$ converges towards the set $\mathcal{E}$ of correlated equilibria.

This is quite a remarkable result: while each of the two players uses a myopic strategy and pays only limited attention to the behavior of the other player, a strong correlation between their behaviors is obtained in the limit.

Definition and interest of the internal regret

A straightforward observation is that in the definition of $\mathcal{E}$ the attention can be restricted to departure functions φ and ψ that only differ from the identity in one point; therefore, $\mathcal{E}$ can be rewritten in a somewhat simpler way as

$$\mathcal{E} = \left\{ \pi \in \Delta(\mathcal{A}\times\mathcal{B}) : \quad \forall (i,k) \in \mathcal{A}^2, \ \sum_{j\in\mathcal{B}} \pi(i,j)\, r(i,j) \geqslant \sum_{j\in\mathcal{B}} \pi(i,j)\, r(k,j) \right.$$
$$\left. \text{and} \quad \forall (j,\ell) \in \mathcal{B}^2, \ \sum_{i\in\mathcal{A}} \pi(i,j)\, s(i,j) \geqslant \sum_{i\in\mathcal{A}} \pi(i,j)\, s(i,\ell) \right\}.$$

Based on this observation, we could define the (per-round) internal regret of the strategy of player A as

$$\max_{(i,k)\in\mathcal{A}^2} \ \frac{1}{T}\sum_{t=1}^{T} (r(k,J_t) - r(i,J_t))\mathbb{I}_{\{I_t=i\}}\,; \tag{2.15}$$

but the Hoeffding–Azuma inequality together with the Borel–Cantelli lemma shows that the asymptotic behavior of the previous quantity is the same as the one of the following quantity, which is slightly simpler to minimize:

$$\overline{R}_T^{\text{int}} = \max_{(i,k)\in\mathcal{A}^2} \ \frac{1}{T}\sum_{t=1}^{T} p_{i,t}\big(r(k,J_t) - r(i,J_t)\big)\,.$$

The latter quantity defines the (per-round) internal regret of player A. The per-round internal regret of player B is defined symmetrically as

$$\overline{S}_T^{\text{int}} = \max_{(j,\ell)\in\mathcal{B}^2} \ \frac{1}{T}\sum_{t=1}^{T} q_{j,t}\big(s(I_t,\ell) - s(I_t,j)\big)\,.$$

We show in the rest of this section how the players can minimize their internal regrets, that is, we exhibit strategies such that

$$\limsup_{T\to\infty} \overline{R}_T^{\text{int}} \leqslant 0 \quad \text{a.s.} \qquad \text{and} \qquad \limsup_{T\to\infty} \overline{S}_T^{\text{int}} \leqslant 0 \quad \text{a.s.}$$

A straightforward adaptation of the proof techniques leading to Propositions 2.3 and 2.7 yields to the following convergence result.

Proposition 2.15. When both players minimize their internal regrets, the sequence $(\overline{\pi}_T)$ of the empirical distributions of action profiles converges almost surely towards the set $\mathcal{E}$ of correlated equilibria.

Automatic conversion of external-regret minimizing strategies into internal-regret minimizing strategies [1]

The original notion of regret defined in Section 2.1 is henceforth referred to as the external regret. In [1] we propose a reinterpretation of the first known internal-regret minimizing strategy, exhibited by [FV99], as a strategy minimizing some external regret.

For all probability distributions $\boldsymbol{p}$ over $\mathcal{A}$ and all pairs $(i,k) \in \mathcal{A}^2$ with $i \neq k$, we denote by $\boldsymbol{p}^{i\to k}$ the image of $\boldsymbol{p}$ by the departure function $\varphi_{i\to k} : \mathcal{A} \to \mathcal{A}$ that only differs from the identity at i, where $\varphi(i) = k$. Put differently, the probability distributions $\boldsymbol{p}^{i\to k}$ and $\boldsymbol{p}$ only differ in the probability masses associated with i and k, which are respectively equal to 0 and p_i for i on the one hand and $p_i + p_k$ and p_k for k on the other hand.

The key observation is then that the internal regret can be rewritten as the (external) regret with respect to the modifications of the strategy of player A parameterized by the $\varphi_{i\to k}$:

$$\overline{R}_T^{\mathrm{int}} = \max_{i\neq k} \sum_{t=1}^{T} r\big(\boldsymbol{p}_t^{i\to k},\, J_t\big) - \sum_{t=1}^{T} r(\boldsymbol{p}_t, J_t)\,.$$

Besides, this rewriting explains the etymology of the notion of internal regret: the comparison class in the definition of the regret is no longer intrinsic and external to the strategy of the player but depends on the contrary on this strategy.

Things are thus equivalent here to player A having meta-actions indexed by the $\varphi_{i\to k}$. To minimize the corresponding external regret it suffices to choose at each round t a distribution $\boldsymbol{p}_t$ satisfying the fixed-point equation

$$\boldsymbol{p}_t = \sum_{i\neq k} \frac{\exp\Big(\eta_t \sum_{s=1}^{t-1} r\big(\boldsymbol{p}_s^{i\to k},\, J_s\big)\Big)}{\sum_{i'\neq k'} \exp\Big(\eta_t \sum_{s=1}^{t-1} r\big(\boldsymbol{p}_s^{i'\to k'},\, J_s\big)\Big)}\, \boldsymbol{p}_t^{i\to k}\,, \tag{2.16}$$

which is obtained by mimicking (2.2) and considering the same learning rates (η_t) as therein. Such a distribution $\boldsymbol{p}_t$ always exists as can be seen by identifying it with a stationary probability distribution of a certain finite Markov chain. The results recalled after (2.2) then show that the per-round internal of this strategy is bounded with probability 1 by

$$\overline{R}_T^{\mathrm{int}} \leqslant \|r\|_\infty \sqrt{\frac{2}{T} \ln\big(N(N-1)\big)}\,.$$

Remark in passing. The argument above can be generalized to convert any external-regret minimizing strategy (not only the exponentially weighted average strategy with time-varying learning rates) into an internal-regret minimizing strategy; it is even applicable in the setting of sequential convex aggregation. Another such conversion trick (only applicable in the setting of randomized prediction, though) was proposed independently by [BM07].

2.4.2 Extension to games with convex and compact action sets $\mathcal{A}$ and $\mathcal{B}$ [4]

We consider in the rest of this section the case where the action sets $\mathcal{A}$ and $\mathcal{B}$ are general topological spaces, each equipped with its Borel σ–algebra, and where the payoff functions $r : \mathcal{A} \times \mathcal{B} \to \mathbb{R}$ and $s : \mathcal{A} \times \mathcal{B} \to \mathbb{R}$ are measurable functions. We still denote by I_t and J_t the respective actions picked by the players at each round t, possibly at random according to probability distributions $\boldsymbol{p}_t$ and $\boldsymbol{q}_t$ over $\mathcal{A}$ and $\mathcal{B}$.

Extension of the definition of correlated equilibrium

We first recall the extension of the definition of correlated equilibria of a finite game (see Definition 2.14) to the more general case considered here; this extension was formulated by [HS89]. To that end we denote respectively by $\mathcal{L}^0(\mathcal{A})$ and $\mathcal{L}^0(\mathcal{B})$ the sets of all measurable functions $\mathcal{A} \to \mathcal{A}$ and $\mathcal{B} \to \mathcal{B}$.

Definition 2.16. The set of correlated equilibria of a two-player game in which the action set of each player is given by a topological space equipped with its Borel σ–algebra is defined as the following (non empty) set of joint distributions:

$$\mathcal{E} = \Bigg\{ \pi \in \Delta(\mathcal{A} \times \mathcal{B}) : \quad \forall \varphi \in \mathcal{L}^0(\mathcal{A}), \quad \mathbb{E}_\pi\big[r(I,J)\big] \geqslant \mathbb{E}_\pi\Big[r\big(\varphi(I),J\big)\Big] \\ \text{and} \quad \forall \psi \in \mathcal{L}^0(\mathcal{B}), \quad \mathbb{E}_\pi\big[s(I,J)\big] \geqslant \mathbb{E}_\pi\Big[s\big(I,\psi(J)\big)\Big] \Bigg\}, \quad (2.17)$$

where the notation $\mathbb{E}_\pi$ indicates that the random vector (I,J) with values in $\mathcal{A} \times \mathcal{B}$ is distributed according to π.

At least at first sight, $\mathcal{E}$ is in general a subset of $\Delta(\mathcal{A} \times \mathcal{B})$ defined by uncountably many constraints while reasonable notions of internal regret (i.e., which can be minimized easily enough by suitable strategies) are built on at most countably many constraints. But under suitable regularity conditions stated in the lemma below, the set $\mathcal{E}$ can be equivalently defined by at most countably many constraints. We denote by $\mathcal{C}(\mathcal{A})$ and $\mathcal{C}(\mathcal{B})$ the sets of all continuous functions $\mathcal{A} \to \mathcal{A}$ and $\mathcal{B} \to \mathcal{B}$.

Lemma 2.17. When the action sets $\mathcal{A}$ and $\mathcal{B}$ are convex and compact subsets of some normed vector spaces, the vector subspaces $\mathcal{C}(\mathcal{A})$ and $\mathcal{C}(\mathcal{B})$, equipped each with the supremum norm, are separable. Let $\mathcal{D}(\mathcal{A})$ and $\mathcal{D}(\mathcal{B})$ be two respective countable dense subsets in $\mathcal{C}(\mathcal{A})$ and $\mathcal{C}(\mathcal{B})$. When the payoff functions r and s are moreover continuous, the set of correlated equilibria $\mathcal{E}$ can then be defined equivalently by considering only in (2.17) the departure functions $\varphi \in \mathcal{D}(\mathcal{A})$ and $\psi \in \mathcal{D}(\mathcal{B})$.

This lemma is proved in two steps. First an ad-hoc version of Luzin's theorem –which states that every measurable function is a continuous function except on a set of small Lebesgue measure– shows that it suffices to consider all continuous functions in (2.17). Next, a dominated-convergence argument guarantees that one can even restrict the attention to the dense subsets $\mathcal{D}(\mathcal{A})$ and $\mathcal{D}(\mathcal{B})$.

Minimization of a generalized internal regret

Convergence towards the set of correlated equilibria. We enforce the assumptions of Lemma 2.17 in the rest of this section. Player A is said to minimize his internal regret when he can ensure that in the limit none of the departure functions in a countable dense subset $\mathcal{D}(\mathcal{A})$ in $\mathcal{C}(\mathcal{A})$ provides a profitable deviation, that is,

$$\forall \varphi \in \mathcal{D}(\mathcal{A}), \qquad \liminf_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T}\Big(r(I_t, J_t) - r\big(\varphi(I_t), J_t\big)\Big) \geqslant 0\,. \tag{2.18}$$

A similar definition holds for player B. Here –unless extra regularity and topological assumptions are made, see [4, Section 5]– no uniform minimization can be achieved in general and in particular no convergence rates results hold.

But the convergence itself is still guaranteed as asserted by the lemma below. Its proof is based on Prohorov's lemma, which states that when $\mathcal{A}$ and $\mathcal{B}$ (and hence, $\mathcal{A}\times\mathcal{B}$) are compact metric spaces, the set $\Delta(\mathcal{A}\times\mathcal{B})$ of all joint distributions is still a compact metric space for the weak–$\star$ topology. The same arguments that led to Propositions 2.3 and 2.7 (the fact that the set of equilibria is defined by closed constraints, the use of sequential compactness properties, and a proof by contradiction) then entail the following result.

Proposition 2.18. Under the assumptions of Lemma 2.17, when both players minimize their internal regrets, the sequence $(\overline{\pi}_T)$ of the empirical distributions of action profiles converges almost surely towards the set $\mathcal{E}$ of correlated equilibria.

Fixed-point strategies to perform the minimization. We now only need to indicate how each player can minimize his internal regret; we consider player A, for instance, and still enforce the assumptions of Lemma 2.17. What follows is a slight simplification of the presentation in [4], which was concerned with families of strategies defined by potential functions and had to resort to the doubling trick. Here, we restricted our attention to the exponential potential throughout this chapter, which is able to deal with a countable number of experts (departure functions) without the need of a doubling trick.

To that end we fix a probability distribution $\mu = (\mu_\varphi)_{\varphi\in\mathcal{D}(\mathcal{A})}$ over $\mathcal{D}(\mathcal{A})$ and resort to the strategy that at each round t chooses a probability distribution $\boldsymbol{p}_t \in \Delta(\mathcal{A})$ which satisfies the fixed-point equation

$$\boldsymbol{p}_t = \sum_{\varphi\in\mathcal{D}(\mathcal{A})} \frac{\mu_\varphi \,\exp\Big(\eta_t \sum_{s=1}^{t-1} r\big(\boldsymbol{p}_s^\varphi,\, J_s\big)\Big)}{\sum_{\phi\in\mathcal{D}(\mathcal{A})} \mu_\phi \,\exp\Big(\eta_t \sum_{s=1}^{t-1} r\big(\boldsymbol{p}_s^\phi,\, J_s\big)\Big)}\; \boldsymbol{p}_t^\varphi\,, \tag{2.19}$$

where for each probability distributions $\boldsymbol{p}$ over $\mathcal{A}$ and each departure function $\varphi \in \mathcal{L}^0(\mathcal{A})$, the image distribution of $\boldsymbol{p}$ by φ, that is, the law of $\varphi(I)$ when the random variable I follows the law $\boldsymbol{p}$, is denoted by $\boldsymbol{p}^\varphi$.

It only remains to see that such a fixed point always exists. To that end, we equip $\Delta(\mathcal{A})$ with its weak-$\star$ topology, which is metrizable; all mappings $\boldsymbol{p} \mapsto \boldsymbol{p}^\varphi$ are continuous

for this topology, so that the left-hand side of (2.19) is a continuous function of $\boldsymbol{p}_t$. The existence of the fixed point of interest is then ensured by the Schauder–Cauty fixed-point theorem, which we recall in its most general and most accomplished statement.

Theorem 2.19 (Schauder–Cauty fixed-point theorem [Cau01]). *Let C be a non-empty convex and compact subset of a topological Hausdorff vector space. Then each continuous map $T : C \to C$ has a fixed point.*

The results around (2.2) can be adapted straightforwardly enough to the case of a countable number of actions to show that the strategy picking probability distributions satisfying the equations (2.19) ensures with probability 1 the following (non-uniform) lower bound:

$$\forall \varphi \in \mathcal{D}(\mathcal{A}), \qquad \frac{1}{T} \sum_{t=1}^{T} \Big(r(\boldsymbol{p}_t, J_t) - r(\boldsymbol{p}_t^{\varphi}, J_t) \Big) \geqslant \Gamma_{\varphi} \sqrt{T} \,,$$

where the constant Γ_{φ} depends on each φ. A joint application of the Hoeffding–Azuma and Borel–Cantelli lemmas concludes the proof that player A indeed minimizes his internal regret in the sense of (2.18).

2.4.3 Extensions to finite games with partial monitoring

This section is motivated by the proof of Theorem 2.10; we indicated after the statement of the latter that it relies on strategies minimizing their internal regrets in some finite games with partial monitoring. We present briefly the fundamental ideas behind the construction of such strategies. To do so, we get back to the notation of Section 2.2.

Review of the (most important) results in [LS07, Per09a, Per09b, Per09c]

Since the pessimistic payoff function ρ is concave (and in general is not linear) in its argument $\boldsymbol{p} \in \Delta(\mathcal{A})$, the best reply to a vector $\underline{h} \in \mathcal{V}$ is (in general) a probability distribution over $\Delta(\mathcal{A})$ whose support is not reduced to a single action. Based on this observation the mentioned articles consider the mixed extension of the original finite game, where at each round the players pick respective probability distributions $\boldsymbol{p}_t \in \Delta(\mathcal{A})$ and $\boldsymbol{q}_t \in \Delta(\mathcal{B})$, now called mixed actions; but instead of drawing actions I_t and J_t at random according to these distributions as in the original game they obtain directly the respective payoffs $r(\boldsymbol{p}_t, \boldsymbol{q}_t)$ and $s(\boldsymbol{p}_t, \boldsymbol{q}_t)$. Of course, a simple application of concentration results shows that the per-round payoffs obtained in the original game and in its mixed extension are asymptotically equal.

The definition (2.15) of the internal regret for finite games is then extended as follows to their mixed extensions. By symmetry we only detail it for player A. We assume that the latter can only pick distributions within a given finite subset $\{\boldsymbol{p}^f,\ f \in \mathcal{F}\}$ of $\Delta(\mathcal{A})$, which is a strong restriction at first sight but seems however reasonable whenever, for

instance, this subset forms a thin enough grid of $\Delta(\mathcal{A})$. For all mixed actions $\boldsymbol{p}^f$ played at least once in the first T rounds, we denote by

$$\overline{\boldsymbol{q}}_T(\boldsymbol{p}^f) = \frac{1}{\sum_{t=1}^T \mathbb{I}_{\{\boldsymbol{p}_t=\boldsymbol{p}^f\}}} \sum_{t=1}^T \boldsymbol{q}_t \, \mathbb{I}_{\{\boldsymbol{p}_t=\boldsymbol{p}^f\}}$$

the empirical average of the mixed actions picked by player B on the rounds when player A resorted to the mixed action $\boldsymbol{p}^f$. Minimizing the internal regret then consists of ensuring that no consistant replacement of a mixed action of player A by another one –the mixed actions of player B being fixed– improves significantly the per-round payoff; that is, we aim at ensuring that for all strategies τ of player B, for all mixed actions $f \in \mathcal{F}$,

$$\liminf_{T\to\infty} \left(\frac{1}{T} \sum_{t=1}^T \mathbb{I}_{\{\boldsymbol{p}_t=\boldsymbol{p}^f\}} \right) \left(\max_{\boldsymbol{p}\in\Delta(\mathcal{A})} \rho\Big(\boldsymbol{p},\, H\Big(\cdot\,, \overline{\boldsymbol{q}}_T(\boldsymbol{p}^f)\Big)\Big) - \rho\Big(\boldsymbol{p}^f,\, H\Big(\cdot\,, \overline{\boldsymbol{q}}_T(\boldsymbol{p}^f)\Big)\Big)\right) \geqslant 0$$

almost surely.

[LS07, Per09b, Per09c] propose strategies minimizing the internal regret defined above; the constructions of [Per09b, Per09c] are based on auxiliary calibrated strategies. The direct implication of Theorem 2.10 is then proved in [Per09a] thanks to such internal-regret minimizing strategies.

Criticisms and perspectives

On the definition of internal regret. The extension of the notion of internal regret is performed by considering a criterion similar to (2.15) up to the replacement of the single actions by mixed actions. To do so an important restriction on the behavior of player A needs to be enforced: he can only pick finitely many different actions during the course of the repeated game. In addition, in the case of sufficient feedback the extended definition does not necessarily coincide with (2.15) while we showed in [3] how to minimize the internal regret defined in (2.15) in this case. There actually exists in all these cases a natural (almost canonical) finite set of mixed actions to be played: [Per09c] proved –as already mentioned earlier in this chapter– that there always exists a finite set that contains, for each vector $\underline{h} \in \mathcal{V}$, a best reply against $\underline{h}$ in the sense of ρ. For instance, in the case of sufficient feedback, this subset is formed by the Dirac masses on the actions in $\mathcal{A}$.

However, a main objection remains: the link between the minimization of the internal regret and the convergence towards some set of equilibria (e.g., the set of correlated equilibria of the mixed extension) is not established here. In contrast, the definitions of the regrets introduced in the previous sections heavily relied on convergence results: we could only provide the right notions of regret to minimize by studying the structure and the defining constraints of the sets of equilibria at hand. (The reader may compare on the one hand (2.8) and (2.9) for the external regret in games with partial monitoring

and on the other hand (2.18) and Lemma 2.17 for the internal regret in games with convex and compact action sets.)

The definition and the study of the possible set of limit equilibria are not straightforward; we recall in passing that [4, footnote 3] shows the identity between the set of correlated equilibria of a finite game and the one of its mixed extension, in the sense that there exists a canonical surjection from the latter onto the former.

Notes about the (memory and computational) complexities. The strategies proposed by [LS07, Per09b] have direct implementations with complexities exponential in T. On the contrary the strategy designed in [Per09c] is efficient (as far as the orders of magnitude in T are concerned); it is based on some common ingredients with the efficient strategy achieving the optimal rates of convergence for the external regret in games with partial monitoring, which we alluded at in Section 2.2.4.

However, this does not imply that an efficient or natural strategy is associated with Theorem 2.10 –in contrast to Theorem 2.1– as the construction of the strategy in [Per09a] relies on the minimization of an internal regret with respect to a grid of an exponential size (in some discretization parameter other than T). An interesting open question would be to design a more satisfactory and more natural approachability strategy associated with Theorem 2.10.

2.5 Perspectives for future research

In this chapter we stated open problems at the end of each of the three main sections, after presenting our contributions and discussing them (see Subsections 2.2.4, 2.3.3, and 2.4.3). All these open problems are rather short-term projects that would sharpen the comprehension or develop the appreciation of existing results. They are to be tackled with Shie Mannor and Vianney Perchet.

CHAPTER 3

Methodological advances in sequential aggregation of experts and empirical studies of their performance

Introduction. New theoretical results in prediction of individual sequences are often assessed on artificial data. In fact, to the best of our knowledge, only few studies on real data were conducted up to now. This is in contrast with the highly general meta-statistical framework of sequential aggregation of experts forecasts introduced in Chapter 1; this framework can indeed be applied to all practical situations in which several experts may be constructed –that is, to virtually all problems of sequential prediction.

The first series of such empirical studies dealt with sequential investment in the stock market and was initiated by [Cov91]; in this chapter we review the achieved contributions in this field. A second –more recent– series considers sports bets, see [DMP+06] or [VZ08]. The forecasts of the experts are given by the odds on the various competitors provided by a set of bookmakers or generated by the bets of the users of a sports-betting website.

The applications on real data that we performed are the following ones: the prediction of electricity consumption on the one hand, where Goude [Gou08a, Gou08b] first illustrated the interest of sequential aggregation methods, and the prediction of air quality on the other hand, for which Mallet and Sportisse [MS06] provided a set of experts and conducted a preliminary study of the performance of simple aggregation strategies.

Table of contents

3.1 Summary of the methodological advances [10]

This chapter is based on the survey article [10] (in French), which summarizes some methodological advances in prediction of individual sequences stated in passing in the articles [7, 13]. This is because the former article is intended to mathematicians and

computer scientists while the latter ones were written for specialists of a given field, e.g., atmospheric sciences. We consider the framework and the notation of Section 1.1.2 to describe these methodological advances.

3.1.1 Practical empirical online tuning of the parameters

Theoretical tunings are too cautious. As we will illustrate on real data the theoretical optimal values of η (the ones that minimize the theoretical bounds) usually exhibit bad practical performance; this is the case for the off-line tuning proposed in Theorem 1.7 or the online tuning described in Section 1.4. An explanation of these disappointing results is that the theoretical bounds are with respect to all individual sequences and thus correspond to too cautious algorithms, which have too long a reaction time. A natural idea therefore consists of increasing the values of the learning rates η_t, the question at hand being to find a powerful and adaptive way to do so.

But theory is useful despite all! Before proceeding, note that these remarks about a desired faster learning do not call into question the general methodology of sequential aggregation of experts forecasts. The applications to real data will indeed show that aggregation strategies tuned with fast enough learning rates perform much better than the best single expert and even than the best constant convex combination of the experts forecasts –while, in addition, both are only known at the end of the forecasting period. This is illustrated by the weight vectors $\boldsymbol{q}_t$ chosen round after round: they absolutely neither look like Dirac masses nor are all similar to a given convex weight vector.

Sequential tuning with the best parameter in the past

We describe the method on an exemple, namely, the family $\mathcal{E}_\eta^{\mathrm{grad}}$ of strategies performing exponentially weighted averages of the cumulative gradients of the losses, which was described in Figure 1.3. The tuning method relies on all the strategies of this family (as η varies) and this is why it is called a tuned meta-strategy (here, based on the $\mathcal{E}_\eta^{\mathrm{grad}}$). To state it formally we now need to write explicitly the dependency of the weight vector $\boldsymbol{p}_t$ on the strategy $\mathcal{E}_\eta^{\mathrm{grad}}$ that prescribes it, which we do by denoting it by $\boldsymbol{p}_t(\mathcal{E}_\eta^{\mathrm{grad}})$.

Statement. The tuned meta-strategy chooses at each time instance t the parameter η_t whose associated strategy obtained the best performance in the first $t-1$ time instances and then resorts to the weight vector $\boldsymbol{p}_t(\mathcal{E}_{\eta_t}^{\mathrm{grad}})$. Formally,

$$\eta_t \in \operatorname*{arg\,min}_{\eta>0} \ \widehat{L}_{t-1}(\mathcal{E}_\eta^{\mathrm{grad}}) \,. \tag{3.1}$$

No theoretical guarantee yet. This online tuning achieves in practice the performance that was intuitively expected, namely, its cumulative loss up to time instance T is often close to the one of $\mathcal{E}_{\eta_T}^{\mathrm{grad}}$, which is by definition the best strategy within the family $\mathcal{E}_\eta^{\mathrm{grad}}$ on the data. We have however no corresponding theoretical guarantee to offer yet.

Issues in the practical implementation. In addition, the computation of the argument of the minimum in (3.1) or even of an ε–minimum is tricky. A simplifying idea is to restrict the minimization to a finite grid of points within the set $\mathbb{R}_+^\star$ of all parameters, where the grid can be constructed or modified online. We will often use logarithmically evenly spaced points between two endpoints. Our empirical studies show that the discretization step has not too strong an influence over the performance, in contrast to the values of the endpoints. [13, 10] propose (without implementing it) a procedure to set these values adaptively over time according to the performance obtained in the past by the previously considered grids. The validation on real data of this procedure needs however to be performed.

3.1.2 Two generic variants of the strategies based on cumulative losses

A frequently overheard criticism. Practitioners often have the following criticism against the strategies described in the previous chapters: they care too much about the far away past since the cumulative losses used at time instance t to compute the weight vectors give the same importance to recent losses and to remote-past losses. The intuition suggests that the knowledge of the most recent past is useful while the one of the remote past seems less profitable (in particular to practitioners familiar with stochastic frameworks, especially when the latter rely on stationary distributions). This is another illustration of the cautious prediction behavior of our robust strategies.

Windowing of the losses: a method with no theoretical guarantee

This is the first generic variant, it discards the past losses as follows. We fix an integer parameter H and only consider the losses obtained in the last H time instances to compute the cumulative losses. For instance, for the strategies $\mathcal{E}_\eta^{\mathrm{grad}}$, this variant consists of replacing the definitions of the components of $\boldsymbol{p}_t$ provided in Figure 1.3 by

$$p_{j,t} = \frac{\exp\left(-\eta \sum_{s=\max\{1,t-H\}}^{t-1} \widetilde{\ell}_{j,s}\right)}{\sum_{k=1}^{N} \exp\left(-\eta \sum_{s=\max\{1,t-H\}}^{t-1} \widetilde{\ell}_{k,s}\right)}$$

for all $t \geqslant 2$ and $j = 1, \ldots, N$.

Unfortunately it seems unlikely that the existence of uniform regret bounds with respect to all individual sequences of experts forecasts and observations be preserved by windowing.

Reconciliation of the viewpoints: discounted losses

It suffices to consider that the importance of the losses is in proportion of their distances to the present: more recent losses are more significant but the remote past still counts (a little). This is made formal by discounting past losses by a positive multiplicative factor that is smaller as the past is less recent.

Statement. We illustrate again the method on the family of strategies $\mathcal{E}_\eta^{\mathrm{grad}}$. It is parameterized by two non-increasing sequences of positive numbers, the discount factors (β_t) and the learning rates (η_t). The definitions of the components of $\boldsymbol{p}_t$ provided in Figure 1.3 are replaced by

$$p_{j,t} = \frac{\exp\left(-\eta_t \sum_{s=1}^{t-1}(1+\beta_{t-s})\,\widetilde{\ell}_{j,s}\right)}{\sum_{k=1}^{N}\exp\left(-\eta_t \sum_{s=1}^{t-1}(1+\beta_{t-s})\,\widetilde{\ell}_{k,s}\right)} \tag{3.2}$$

for all $t \geqslant 2$ and $j = 1, \ldots, N$.

Existence of theoretical guarantees on the performance. We could prove uniform regret bounds for the instantiation above of this second variant, these bounds depending, of course, on the sequences (η_t) and (β_t). The latter sequence should, for instance, not decrease too quickly towards 0.

Theorem 3.1. *The exponentially weighted average strategy with discounted losses introduced in (3.2) minimizes the regret in the sense of Aim 1.1 whenever the learning rates and the discount factors satisfy*

$$t\,\eta_t \to \infty \qquad \text{and} \qquad \eta_t \sum_{s=1}^{t-1} \beta_s \longrightarrow 0$$

as $t \to \infty$ and Assumption 1.6 is satisfied.

Proof. A detailed proof can be found in the technical report [MMS07, Chapter 6]. The idea is to consider an approximation scheme in which the –not too large– discrepancies between the weight vectors with and without discounting –defined respectively in (3.2) and in Figure 1.3 up to the replacement of the fixed learning rates by online values η_t– are quantified. These discrepancies are then added to the regret bound of an adaptive version of the strategy of Theorem 1.7 (alluded at in the comments following the statement of the theorem). □

Literature review. Discounting factors were introduced in prediction of individual sequences by [CBL06, Section 2.11]; but it was crucial for the analysis carried therein that the number of time instances T be fixed and known beforehand, which is not a restriction that we could consider here. In game theory discount factors model in particular the interest rates and give consequently more importance to less recent payoffs.

3.1.3 Sequential linear regressions with regularization factors

The framework of sequential linear regression. In this section we restrict our attention to the case where the observation and prediction sets are given by the real line, $\mathcal{Y} = \mathcal{X} = \mathbb{R}$,

and where the loss function ℓ is the quadratic loss, $\ell(x,y)=(x-y)^2$. The statistician is now allowed to pick at each time instance arbitrary weight vectors $\boldsymbol{u}_t=(u_{1,t},\dots,u_{N,t})$ in $\mathbb{R}^N$; no nonnegativity or summation-to-1 constraints are enforced anymore. The forecasts output by the statistician are then given by the linear combinations

$$\widehat{y}_t=\sum_{j=1}^{N}u_{j,t}f_{j,t}\,. \tag{3.3}$$

Possible reduction of the bias but lack of interpretation. The advantage of this framework is that by removing the constraint that the weights should sum up to 1 we now allow the aggregation strategies to compensate some bias that would be common to all experts. We actually observe such a compensation for some data sets, on which the obtained weights sum up to a quantity slightly smaller than 1, like 0.99. The drawback is however that in general the weight vectors $\boldsymbol{u}_t$ have many negative components, which makes them less interpretable than their convex counterparts $\boldsymbol{p}_t$.

ℓ^2–regularization factors: the ridge regression forecaster

The first strategy of interest is based on the ridge regression, which was introduced by [HK70] in a stochastic context and later imported by [AW01] and [Vov01] into the setting of prediction of individual sequences.

Statement. This strategy relies on a ℓ^2 regularization; to state it we denote by

$$\|\boldsymbol{u}\|_2=\sqrt{\sum_{j=1}^{N}u_j^2}$$

the Euclidian norm of a vector $\boldsymbol{u}\in\mathbb{R}^N$. The ridge regression forecaster is parameterized by a parameter $\lambda>0$ and will be referred to as $\mathcal{R}_\lambda$. It chooses at each time instance $t\geqslant 1$ a weight vector $\boldsymbol{u}_t$ satisfying

$$\boldsymbol{u}_t\in\operatorname*{arg\,min}_{\boldsymbol{v}\in\mathbb{R}^N}\left\{\lambda\|\boldsymbol{v}\|_2^2+\sum_{s=1}^{t-1}\left(y_s-\sum_{j=1}^{N}v_jf_{j,s}\right)^2\right\} \tag{3.4}$$

with the convention that a sum over no element is null (so that, for instance, $\boldsymbol{u}_1$ is the null vector).

Theoretical performance bound. To state it in a compact way we define the (line) vector $\boldsymbol{f}_t=(f_{1,t},\dots,f_{N,t})$ of the experts forecasts at time instance t.

Theorem 3.2 (see [CBL06, Section 11.7]). *Consider the matrix $M_T=\sum_{t=1}^{T}\boldsymbol{f}_t^{\mathrm{T}}\boldsymbol{f}_t$ and denote by $\mu_{1,T},\dots,\mu_{N,T}$ its eigenvalues. For all sequences of experts forecasts*

$\boldsymbol{f}_1, \ldots, \boldsymbol{f}_T$ and all sequences of observations $y_1, \ldots, y_T$, the regret of $\mathcal{R}_\lambda$ with respect to each weight vector $\boldsymbol{v} \in \mathbb{R}^N$ is bounded from above as follows:

$$\sum_{t=1}^{T} \left(y_t - \sum_{j=1}^{N} u_{j,t} f_{j,t} \right)^2 - \sum_{t=1}^{T} \left(y_t - \sum_{j=1}^{N} v_j f_{j,t} \right)^2 \leqslant \frac{\lambda}{2} \|\boldsymbol{v}\|_2^2 + \left(\sum_{j=1}^{N} \ln\left(1 + \frac{\mu_{j,T}}{\lambda}\right) \right) \max_{t \leqslant T} \left(y_t - \sum_{j=1}^{N} u_{j,t} f_{j,t} \right)^2 .$$

Remarks and comments. To drop all dependencies of the bound in the sequences of experts forecasts $f_{j,t}$ and observations y_t it suffices to restrict $\mathcal{X}$ and $\mathcal{Y}$ to a bounded domain $[-B, B]$ (at the cost, however, of a deterioration of the orders of magnitude of the regret bound; details are omitted). We end this paragraph by underlining that here again, a tuning issue arises for λ. The computation of its theoretical optimal value with respect to the bound of Theorem 3.2 is tricky; in practice, we advise the reader to resort to the sequential data-based tuning proposed in Section 3.1.1.

ℓ^1–regularization factors: the sequential Lasso forecaster

We recently introduced the following other variant of the sequential linear regression, which is in the spirit of some modern regularized regression methods of the stochastic case.

Statement. We replace the ℓ^2–regularization factor used in (3.4) by an ℓ^1–regularization: to that end we denote by

$$\|\boldsymbol{u}\|_1 = \sum_{j=1} |u_j|$$

the ℓ^1–norm of a vector $\boldsymbol{u} \in \mathbb{R}^N$. For a given regularization constant $\lambda > 0$, the sequential Lasso forecaster chooses at each time instance $t \geqslant 1$ a weight vector

$$\boldsymbol{u}_t \in \operatorname*{arg\,min}_{\boldsymbol{v} \in \mathbb{R}^N} \left\{ \lambda \|\boldsymbol{v}\|_1 + \sum_{s=1}^{t-1} \left(y_s - \sum_{j=1}^{N} v_j f_{j,s} \right)^2 \right\} . \tag{3.5}$$

We denote this strategy by $\mathcal{L}_\lambda$.

Remarks and comments. [Tib96] introduced and studied the Lasso regression in a stochastic setting, with a booming success. It in particular turned out to be remarkably suited to high-dimensional regression problems. Indeed, the advantage of the ℓ^1–regularization in (3.5) is that it leads to weight vectors $\boldsymbol{u}_t$ with only few non-zero components. A drawback is however that these vectors cannot be given in closed form –but there exists algorithms, like the LARS algorithm of [EHJT04], that provide

an efficient computation of their values on the data. This is to be compared to the minimization problem (3.4) for which it is easy to exhibit such a closed form expression of $\boldsymbol{u}_t$ as a function of λ and of past data.

Perspective for future research. To the best of our knowledge there exists no regret bound yet for the sequential Lasso regression with respect to individual sequences; the desired form of the bound would be similar to the one stated in Theorem 3.2.

3.2 Interlude: Outline of the empirical studies

We provide a standardized outline of the treatment of a new data set. Since our aim in this thesis is only to assess *in hindsight* the interest of using sequential aggregation strategies (the operational performance of the latter), we assume that all observations are available. Of course, the ultimate goal is to design fully automatic strategies but doing so we can also study the performance of semi-automatic strategies.

Outline of the empirical studies of performance of the sequential aggregation methods

1. Design some experts.
2. Choose a loss function and evaluate the performance of the experts.
3. For each family of strategies compute the performance corresponding to the best constant choices of the parameters in hindsight.
4. Measure the cost of the automatic tuning and assess the quality of the operational performance.

1. Design some experts. The guideline is to design them so that –as much as possible– they exhibit varied enough behaviors in order that the aggregation strategies have a sufficient flexibility in the output aggregated forecasts. Constructing the experts is usually the responsibility of the partner of the statistician because of his knowledge of the field of application and of the methods –classical and more modern ones– that are likely to exhibit a good performance. These methods can rely on some tuning parameters that were set on data sets anterior to the data set at hand (see, for instance, the construction of the experts for the electricity consumption in Section 3.5.2).

2. Choose a loss function and evaluate the performance of the experts. By evaluation of the performance of the experts we mean the assessment of the accuracy obtained by some simple strategies like the uniform average of the experts forecasts (which is a strategy that is easy to implement online) or by some oracles; this assessment is given by their cumulative losses. By oracles we mean strategies that cannot be defined online and that require the beforehand knowledge of the whole data set: the best single expert, the best constant convex combination of the experts, the best constant linear combination of the experts. Finally, the prescient strategy is the strategy that is only constrained

by outputting at each time instance a convex (or linear) combination of the experts; it indicates a bound on the performance that no aggregation strategy can improve given the data set (given the experts forecasts and the observations).

3. For each family of strategies compute the performance corresponding to the best constant choices of the parameters in hindsight. The aggregation strategies often require the tuning of a small number (usually, one or two) of user parameters. For instance, the family $\mathcal{E}_\eta^{\mathrm{grad}}$ of strategies performing exponentially weighted averages of the cumulative gradients of the losses relies on one parameter η. What we do here is to tabulate the performance on a thin grid of possible parameters and compare the best accuracy obtained in this way to the performance of the reference strategies and oracles of the previous step –with the hope that the aggregation strategies will perform better than the oracles in addition of being implementable online.

4. Measure the cost of the automatic tuning and assess the quality of the operational performance. We then implement the tuned meta-strategy of Section 3.1.1 based on the families considered in the previous step and look how different is its performance with respect to the best of the underlying strategies (the ones computed in the previous step). This is the most crucial step of the empirical study since it indicates the performance that would have been achieved for real by outputting sequentially aggregated forecasts based on the experts constructed in the first step –hence the notion of operational performance.

3.3 Sequential investment in the stock market [1]

The first series of empirical studies on the practical interest of the sequential aggregation techniques was about sequential investment in the stock market and was initiated by [Cov91]. The observations are formed by the daily evolutions of 36 assets of the New-York stock exchange between 1963 and 1985 while the experts are simply identified with each of these assets. Twenty articles at least studied the performance obtained by sequential aggregation strategies on this data set but we only discuss in the rest of this section the results obtained by [HSSW98, BEYG00, GLU06, AHKS06] and by our contribution [1].

Warning! In this section we briefly describe why this simplified framework and its associated data set were criticized by academic researchers as well as by some professional R&D researchers in the finance industry; I had direct conversations with members of both groups (and was also reported additional conversations of colleagues with members of the second group).

Brief overview of the obtained results. [BEYG00] shows that the returns obtained by the aggregation strategies of [Cov91] are close to the ones achieved by the allocation strategy

that redistributes its capital everyday evenly among the 36 assets; the latter strategy corresponds to the constant use of the uniform weight vector. [HSSW98] (by considering the family $\mathcal{E}_\eta^{\text{grad}}$), [1] (by minimizing some internal regret), and later [AHKS06] (by a convex aggregation relying on the optimization of a criterion via Newton's method) gradually improved the financial performance of sequential aggregation strategies coming with theoretical guarantees on their regrets with respect to individual sequences. [GLU06] and some posterior contributions obtain returns larger by several orders of magnitude but resort to do so to strategies exploiting a stochastic assumption on the behavior of the market: that it can be modeled by a stationary process.

Criticisms formulated over time. To the best of our knowledge most of the empirical studies –including ours– do not fully implement the outline described above in Section 3.2 and only consider its steps 2 and 3. In particular they do not design real experts (step 1) and merely identify each asset with an expert, which is the most important criticism. On the contrary they should ask R&D departments of the finance industry to provide them with a set of true base investment strategies and aggregate the investment portfolios recommended by the latter.

On the one hand these studies also tabulate the performance only for a set of constant choices of the tuning parameters and never discuss the operational performance obtained via an automatic sequential tuning (step 4). On the other hand even the non-operational returns of the considered strategies are far from the ones of the best constant convex combination of the assets, which is in strong contrast with the improvements obtained by the former on the latter in other settings, which are presented in the next two sections of this chapter.

To these academic criticisms on the methodology can be added other ones on the data set itself, which suffers from what is known as the "survivor bias" as it only contains assets that did not go bankrupt during the considered period of time. In addition –and maybe most importantly– researchers, see, e.g., [GLU06], quickly realized that the most impressive returns obtained by sequential strategies were essentially linked to two specific assets, "Kin Ark" and "Iroquois", which have cyclic, correlated, and highly foreseeable behaviors.

Finally, the exchanges with the financial industry are sometimes one-sided: the obtained feedback on the performance and on the interest of the sequential aggregation strategies is vague and non informative, the R&D researchers of the finance industry are reluctant to detail their operational constraints, but they are always eager for learning new and original strategies. All in all, it is difficult to design the experts hand in hand, as true partners –while this step 1 of the outline detailed in Section 3.2 is crucial.

Conclusion. This is why I never went back to the problem of sequential aggregation of portfolios after my PhD thesis. I instead preferred working in two other fields, in which the preliminary step consisted of identifying solid and renowned partners that were able to provide experts of good quality.

3.4 Air quality forecasting [7]

We present in this section the context and the data set relative to the problem of air quality forecasting. We briefly explain how to adapt the general strategies presented above and in the previous chapters to this setting and then provide an overview of their performance. All related details can be found in the articles [7, 10], as well as in the technical reports [GMS08, MMS07] on which these articles rely.

3.4.1 Description of the considered data set and of the experts used

The data set at hand corresponds to the time period between April 28 and August 31, 2001, which contains thus $T = 126$ days, and to a geographical localization over France and Germany: 241 sites (also called stations in the sequel) are available –116 in France and 81 in Germany– and they are uniformly distributed over each of the countries. We only discuss in this section the results obtained for the forecasting of daily ozone peaks: with each day t and each site s we associate the quantity y_t^s, which is the maximal value of the ozone concentration during day t at site s. The indexes t and s take respective values in the sets $\{1, \dots, 126\}$ and $\mathcal{N} = \{1, \dots, 241\}$. The measures are given in micrograms per cubic meter ($\mu\mathrm{g\,m}^{-3}$), a unit that will generally be omitted in the sequel. In this respect we recall that typical concentrations are of the order of 40 $\mu\mathrm{g\,m}^{-3}$ to 150 $\mu\mathrm{g\,m}^{-3}$ and that French authorities need respectively to inform and to alert people whenever the concentrations are expected to be above 180 $\mu\mathrm{g\,m}^{-3}$ and 240 $\mu\mathrm{g\,m}^{-3}$.

Missing data. About 30 000 peaks are therefore to be predicted but only about 27 500 of them could be measured within the period of interest (missing values correspond, among others, to temporary failures of some meteorological stations). In the sequel we denote by $\mathcal{N}_t$ the set of stations that are active at day t –so that for all t, only the observations y_t^s with $s \in \mathcal{N}_t$ are available.

Other data sets. Two other data sets are considered in [7]: a Europe-wide and a France-wide ones. The hourly forecasting of the ozone concentration is also studied for the three mentioned data sets.

Construction of the experts

We use $N = 48$ experts constructed in [MS06] and integrated in the modeling system Polyphemus[1]. Each expert results from three choices: a physical formulation (how the chemical species evolve all together); a numerical discretization scheme (since the physical formulation is in terms of partial differential equations, of which an approximate solution needs to be computed); a set of input data (meteorological data and measurements of

[1] See `http://cerea.enpc.fr/polyphemus/`

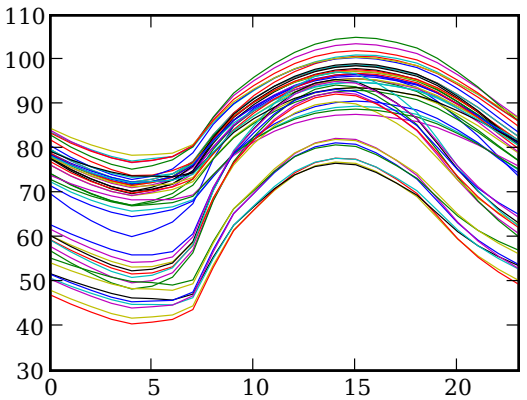

Figure 3.1. Profiles of the forecasts of the ozone concentration (μg m^{-3}) output by the 48 experts, averaged over space and time; x–axis: hours of the day; y–axis: concentrations.

other polluting species). The possible sets of choices are detailed in [MS06, Section 2.2] and lead to 48 experts. In our meta-statistical view, however, experts are simply forecasting black boxes whose accuracies can be improved in some automatic way by aggregation of their forecasts. These experts are indexed by $j \in \{1, \dots, 48\}$ and offer each a forecast $f^s_{j,t}$ for the peak that will occur on day t at site s. They actually even offer fields of forecasts over the entire European continent, that is, forecasts for each point of a thin grid of locations over Europe.

The experts exhibit varied behaviors. Figure 3.1 shows that the experts forecasts are scattered: even the averages of the hourly forecasts over time (over all days of the prediction period) and space differ strongly between the experts, sometimes by a multiplicative factor as large as 2. The shapes of the averages all correspond to the typical concentration profile, with a peak measured at the end of the afternoon and a minimum achieved at the end of the night. But this does absolutely not mean that the experts are given by mere translations of a reference expert: the similarities in the shown profiles are only due to the strong averaging over space and time. As we will illustrate in the sequel the experts indeed have varied behaviors and performance, over space as well as over time.

Aggregation: uniform over time but variable over space

In this section we restrict our attention to aggregation strategies that use the same convex or linear weight vector $\boldsymbol{p}_t$ or $\boldsymbol{u}_t$ to aggregate the experts forecasts at all sites; that is, this vector depends on t but not on s. This constraint can be relaxed to achieve a better overall performance (see [7, Section A.1]) but its advantage is to provide stronger and more interpretable forecasts: the experts provide fields of forecasts and the latter can be combined to form an aggregated field of forecasts, thus providing predictions

even between the sites (though the accuracy of these predictions cannot be evaluated since no observation is available between the sites).

Assessment of the accuracy of the forecasts

Before proceeding we first need to define the loss function used to assess the accuracy of the output forecasts.

Observation and prediction sets $\mathcal{Y}$ and $\mathcal{X}$. The observations at each station lie in the domain $[0, 300] \cup \{\perp\}$, where the symbol $\perp$ denotes a missing value (when the station is inactive) and the upper value 300 is a bound on the maximal ozone concentration that can arise in France or Germany. Similarly, the experts forecasts for a given day and site are assumed to lie in the interval $[0, 300]$. The stations being indexed by the set $\mathcal{N}$, the observation and prediction sets are then given by

$$\mathcal{Y} = \left([0, 300] \cup \{\perp\}\right)^{\mathcal{N}} \qquad \text{and} \qquad \mathcal{X} = [0, 300]^{\mathcal{N}} .$$

Instantaneous losses. The loss function $\ell : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$ equals, for all pairs $\boldsymbol{y} = (y_s)_{s \in \mathcal{N}}$ and $\boldsymbol{x} = (x_s)_{s \in \mathcal{N}}$ of elements in $\mathcal{Y}$ and $\mathcal{X}$,

$$\ell(\boldsymbol{x}, \boldsymbol{y}) = \sum_{s : y_s \neq \perp} (x_s - y_s)^2 .$$

This function satisfies Assumption 1.6 so that we can instantiate in the sequel the strategies $\mathcal{E}_\eta^{\mathrm{grad}}$ of Section 1.2.3.

The instantaneous loss at day t of an aggregation strategy resorting to the (convex or linear) weight vector $\boldsymbol{v}_t = (v_{1,t}, \ldots, v_{N,t})$ to combine the experts forecasts is then equal, with the notation above, to a quantity that for the sake of concision we denote by $\ell_t(\boldsymbol{v}_t)$:

$$\ell_t(\boldsymbol{v}_t) = \sum_{s \in \mathcal{N}_t} \left(y_t^s - \sum_{j=1}^{N} v_{j,t} f_{j,t}^s \right)^2 . \tag{3.6}$$

The notation ℓ_t encompasses both the experts forecasts $f_{j,t}^s$ and the observations y_t^s; by using the formula $\ell_t(\boldsymbol{v}_t)$ we only make explicit the dependencies of the instantaneous losses on the uniform weight vectors $\boldsymbol{v}_t$ output by the aggregation strategies.

Global assessment: root mean squared error. In this section only, since the prediction period is short (it lasts 126 days), the root mean squared error (RMSE) of the experts and of the reference strategies is not computed over the whole prediction period but only on its last 96 days, which leaves 30 days to the strategies as an initial learning period. We denote by $\{t_0, \ldots, T\} = \{31, \ldots, 126\}$ the indexes of the days when the evaluation thus takes place. Formally, given the (linear or convex) weight vectors $\boldsymbol{v}_{t_0}, \ldots, \boldsymbol{v}_T$ chosen by

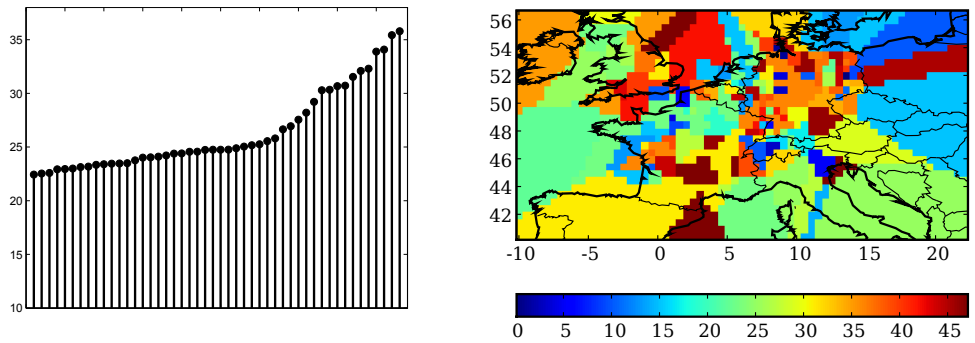

Figure 3.2. Graphical representation of the performance of the experts: root mean squared errors of the experts on the considered data, sorted in increasing order (left) and coloring of the map of Europe based on the index of the best local expert (right).

an aggregation strategy $\mathcal{S}$, its root mean squared error is defined by

$$\text{RMSE}(\mathcal{S}) = \sqrt{\frac{1}{\sum_{t=t_0}^{T} |\mathcal{N}_t|} \sum_{t=t_0}^{T} \ell_t(\boldsymbol{v}_t)}$$

(where $|\mathcal{N}_t|$ denotes the cardinality of $\mathcal{N}_t$).

Ensuring that a strategy has a small regret is equivalent to ensuring that its root mean squared error is close, for instance, to the one of the best expert or of the best constant convex combination of the experts forecasts. In the sequel we will report the results only in terms of the RMSEs instead of the regrets as the former are the performance criterion used by the practitioners and the latter are mostly used a tool to get accurate forecasts via their minimization.

Additional notation. For all sequences $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_T \in \mathbb{R}^N$ of linear weight vectors, we denote by $\boldsymbol{v}_1^T$ the strategy that –independently of the observations and of the experts forecasts– chooses the weight vector $\boldsymbol{v}_t$ at time instance t; its root mean squared error is denoted by $\text{RMSE}(\boldsymbol{v}_1^T)$. When all these weight vectors $\boldsymbol{v}_t$ are equal to some common value $\boldsymbol{v}$, this error is simply denoted by $\text{RMSE}(\boldsymbol{v})$. We also recall that δ_j is referring to the Dirac mass on j, which corresponds to following the forecast of expert j.

3.4.2 Performance of the considered experts and of some reference aggregation strategies

Another illustration of the varied behaviors exhibited by the experts. The stem plot of Figure 3.2 shows the root mean squared errors suffered by the experts on the considered data set; they lie between 22.43 and 35.79. One could think that an expert or a small group of experts outperform clearly all other experts –but this is not the case. The map of Europe shown in Figure 3.2 is colored according to the indexes of the best expert for

Name of the reference strategy	Formula	Value
Uniform average	$\text{RMSE}\big((1/48,\dots,1/48)\big)$	$= 24.41$
Best expert	$\min_{j=1,\dots,48} \text{RMSE}(\delta_j)$	$= 22.43$
Best convex combination	$\min_{\boldsymbol{q}\in\mathcal{P}} \text{RMSE}(\boldsymbol{q})$	$= 21.45$
Best linear combination	$\min_{\boldsymbol{u}\in\mathbb{R}^N} \text{RMSE}(\boldsymbol{u})$	$= 19.24$
Prescient strategy	$\min_{\boldsymbol{u}_1,\dots,\boldsymbol{u}_T\in\mathbb{R}^N} \text{RMSE}\big(\boldsymbol{u}_1^T\big)$	$= 11.99$

Table 3.1. Performance of some reference strategies on the data set of ozone peaks.

each zone of Europe (the expert with the smallest root mean squared error); no expert is uniformly the best one over the whole space and only a notion of best local expert could be defined. We also note that many experts are best local experts for at least one part of the space. This illustrates on the one hand that all experts are useful and provide information and on the other hand that their behaviors and their performance vary over space (we explain below why one can also say that they vary over time).

Performance of some reference strategies. Table 3.1 shows the performance of the reference strategies indicated in Section 3.2.

3.4.3 Performance of some forecasting strategies (by aggregation)

In [7] and [GMS08] we studied about twenty strategies but reproduce only here a brief summary of the performance obtained by three strategies (and variants thereof).

Exponential weighted averages of the gradients of the losses $\mathcal{E}_\eta^{\text{grad}}$ and ridge regression $\mathcal{R}_\lambda$

We discuss in this paragraph the families of strategies $\mathcal{E}_\eta^{\text{grad}}$ (Section 1.2.3) and $\mathcal{R}_\lambda$ (Section 3.1.3). The former family relies on the pseudo-losses (1.11) associated with the gradients of the losses introduced in (3.6) and needs no further comments. In contrast, the study of latter family required the extension of the definition (3.4) and of its associated regret bound to the case where $|\mathcal{N}_t|$ quadratic loss terms (one for each active site) instead of a single one are added at each time instance.

Performance for fixed parameters and operational performance. The first four columns of Table 3.2 show the performance of these two families for various constant values of the parameters (lots of other values than the ones shown were considered). The value 5×10^{-7} approximatively corresponds to the theoretical optimal value $\eta^\star$ recommended by Theorem 1.7 but is far from being the best value in practice.

Value for η	5×10^{-7}	5×10^{-6}	2×10^{-5}	10^{-4}	Grid
RMSE of $\mathcal{E}_\eta^{\text{grad}}$	22.89	21.70	<u>21.47</u>	22.10	21.77

Value for λ	0	100	10^4	10^6	Grid
RMSE of $\mathcal{R}_\lambda$	20.79	<u>20.77</u>	21.13	21.80	20.81

Table 3.2. Performance of the families of strategies $\mathcal{E}_\eta^{\text{grad}}$ and $\mathcal{R}_\lambda$ for various constant values of their parameters η and λ, as well as of the tuned meta-strategies based on them. For each family the smallest RMSE among those obtained for constant parameters is underlined.

This is why –as explained in Section 3.1.1– we resort to an online tuning via a grid of parameters. We use here its simplest version, in which the grid is fixed once for all and does not changer over time. The respective grids for the families $\mathcal{E}_\eta^{\text{grad}}$ and $\mathcal{R}_\lambda$ consist of 11 logarithmically evenly spaced points between 10^{-8} and 10^{-4} on the one hand, 1 and 10^6 on the other hand. The results obtained this way are shown in the last column of Table 3.2. The cost for this automatic tuning is small enough compared to the performance obtained by the choices of the best parameters in hindsight.

Performing aggregation is not following a single expert. We conclude this paragraph by noting that the considered aggregation strategies do not focus in practice on a single expert and that on the contrary the weights associated with the vectors may vary rapidly and significantly over time. This is illustrated by Figure 3.3, where we considered the optimal-in-hindsight parameters η and λ. These variations occur because the performance of the experts themselves change over time; we recall that Section 3.4.2 already underlined that they were varying also over space.

Variants of these two families of strategies: windowing and discounted losses

We consider in this paragraph the generic variants presented in Section 3.1.2 and apply them to the families of strategies studied above; we recall that we only described formally the variants for the case of exponentially weighted average strategies but that they can be extended in a natural way to the case of sequential regressions. Table 3.3 summarizes the results obtained. By original versions therein we mean the versions considered in Table 3.2 and we only report the smallest RMSE that was obtained for constant choices of the parameters. Similarly, the RMSEs indicated in the table for the two variants correspond to an optimization in hindsight of the parameters on the data set (more details are provided in [10] like, for instance, the description of these parameters).

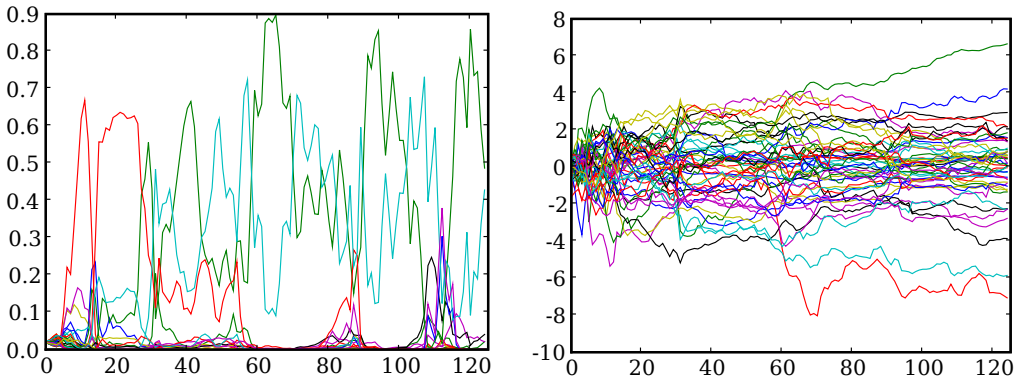

Figure 3.3. Graphical representation of the convex weight vectors chosen by $\mathcal{E}^{\mathrm{grad}}_{2\times 10^{-5}}$ (left) and of the linear weight vectors chosen by $\mathcal{R}_{100}$ (right), according to time.

Family	Original	Windowing	Discounted losses
$\mathcal{E}^{\mathrm{grad}}_{\eta}$	21.47	21.37	21.31
$\mathcal{R}_{\lambda}$	20.77	20.03	19.45

Table 3.3. RMSEs achieved by members of various families of strategies, each tuned with the best parameter(s) in hindsight: original versions, windowing variants, variants relying on discounted losses.

Conclusion. We note that giving a smaller weight to the past –either by windowing or by considering discounted losses– improves the performance, as the practitioners suggested. However, since the results for discounted losses are better than the ones achieved by windowing, we conclude that the remote past should not be totally discarded.

Ridge regression: robustness and automatic reduction of the biases

Robustness. We perform in [7, Sections 4.3.2 and 4.3.3] a robustness study of the best strategy obtained so far, the ridge regression forecaster with discounted losses. We check that the excellent global performance (averaged over all sites and prediction days) does not come at the price of some local disasters.

Automatic reduction of the biases. In addition the ridge regression forecaster (especially its version with discounted losses) can be used to perform some pre-treatment on the experts: in order to reduce their biases. We recall in passing that this ability to reduce the biases was the motivation to resort to linear weight vectors instead of convex ones and was a compensation for the loss of interpretability. Formally we fix an expert k and propose at each time instance t the forecasts $b_t f^s_{k,t}$ instead of the $f^s_{k,t}$, where b_t is a

Expert	original RMSE	RMSE after pre-treatment
Best	22.43	21.66
Reference	24.01	22.43
Worst	35.79	24.78

Table 3.4. Reduction of RMSEs obtained with the pre-treatment consisting of applying to each single expert the optimal ridge regression forecaster with discounted losses studied in Table 3.3.

multiplicative factor given by

$$b_t \in \operatorname*{arg\,min}_{b\in\mathbb{R}} \left\{ \lambda |b|^2 + \sum_{t'=1}^{t-1} (1+\beta_{t-t'}) \sum_{s\in\mathcal{N}_{t'}} \left(y^s_{t'} - b f^s_{j,t'}\right)^2 \right\} .$$

In this case b_t is always nonnegative and is closer to 1 as the original forecasts of expert k have smaller biases: the aim is indeed –as asserted by Theorem 3.2– to perform almost as well as the best of the meta-experts forecasting $b f^s_{j,t}$ at each site s and at each time instance t, where the nonnegative parameters b are multiplicative scaling factors indexing the meta-experts.

Table 3.4 illustrates the interest of this pre-treatment on three experts among the 48: the best and worst ones (i.e., with smallest and largest RMSEs), as well as a reference expert constructed by considering the most common values for the choices described in Section 3.4.1 (see [MS06, Section 2.2] for further details). In all cases a reduction of the RMSE is achieved. An idea that we have not implemented yet would be to apply this pre-treatment to all experts and to aggregate their corrected forecasts instead of the original ones.

Sequential Lasso forecaster with discounted losses

To pave the way for future data sets that would correspond, for instance, to a huge number of experts with respect to the length of the considered time period we performed a preliminary study in [GMS08]. The topic was to perform simultaneously a selection of a small subset of the experts and a linear combination of the forecasts of the selected experts; the small subset is of course to change over time. We resorted to that end to a variant of the sequential Lasso forecaster of Section 3.1.3 obtained by considering discounted losses. After an optimization in hindsight of its parameters (see [10] for the details) this strategy achieved a RMSE of 19.31; the latter value lies between the RMSE of the linear oracle –which is equal to 19.24– and the one of the discounted variant of the ridge regression forecaster –which is equal to 19.45. The simultaneous selection & aggregation performed by this strategy is shown in Figure 3.4. We note that

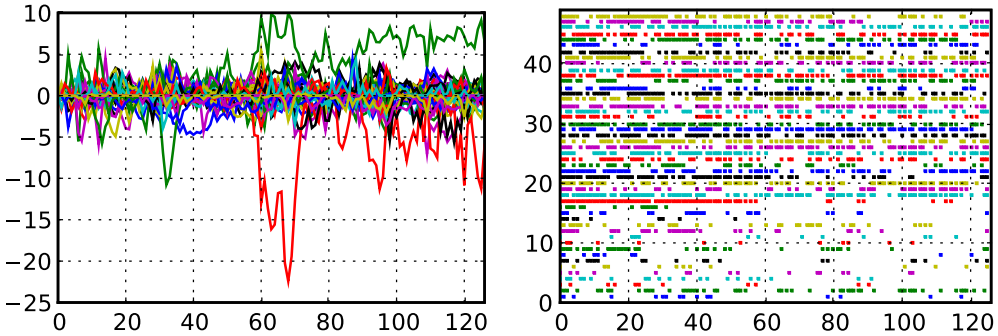

Figure 3.4. Graphical representation of the behavior of the sequential Lasso forecaster with discounted losses tuned with its optimal-in-hindsight parameters: evolution of the chosen linear weight vectors (left) and of the selected experts (right: a square means that the expert has a zero weight).

typically about twenty experts are eliminated at each time instance and that the linear aggregation is performed only over a subset of about thirty experts.

3.5 Forecasting of the electricity consumption [13]

We consider in this section the half-hourly forecasting of the global electricity consumption of the customers of Electricité de France (EDF), the largest electricity provider in France. The results presented here are extracted from the submitted article [13] as well as from the corresponding technical report [DGS09]. Both also study the hourly consumption of the customers of the Slovakian subbranch of EDF.

Specialized experts. The main difficulty –and chance, though– is that the experts of the data set are specialized and only output forecasts in some contexts, hence, irregularly. (Such experts were also called sleeping experts in the literature, see below the review of the latter.) For instance, some experts may be designed to output forecasts expected to be accurate in winter and rather crude in summer; there can be experts dedicated to working days and others dedicated to week-ends and public holidays. Dealing with such experts may be a chance as their specialization probably implies more accurate forecasts on those time instances when they provide some. It is a difficulty however at first sight since the definitions and results of Chapter 1 need to be adapted to this setting.

Outline of the study. We first indicate briefly how to mathematically deal with this setting, present then the data set and in particular the constructed experts, and conclude by describing the performance obtained by the considered aggregation strategies.

3.5.1 How to take advantage of specialized experts

Literature review. To the best of our knowledge this setting was not much considered in the field of prediction of individual sequences. The first references are [Blu97] and [FSSW97]; they respectively introduce and formalize the framework of specialized experts. Two other papers focusing on other topics but mentioning in passing results for the case of specialized experts are [BM07, Sections 6–8] and [CBL03, Section 6.2]. All these references deal with convex aggregation; there seems to be no result yet for linear aggregation (e.g., for an adaptation of the ridge regression forecaster). Preliminary attempts and partial results for the problem of linear aggregation in the context of sleeping experts are provided in the technical report [DGS09] but are not satisfactory and need to be rethought.

Mathematical statement of the problem. We use again the notation of Chapter 1. The prediction set $\mathcal{X}$ is extended with the element $\perp$, which has the following meaning. That expert $j \in \{1, \dots, N\}$ proposes at time instance t the value $f_{j,t} = \perp$ means that the context is not the required one for it to output a forecast (i.e., some external conditions are not met), in which case the expert refrains from forecasting. It is then said inactive. In contrast experts proposing other forecasts in $\mathcal{X}$ are said active. We assume that at each time instance t at least one expert is active and we denote by E_t the non-empty set of these active experts. As indicated above we restrict our attention in this thesis to convex weight vectors. All in all each prediction strategy $\mathcal{S}$ thus chooses at each time instance t a convex weight vector $\boldsymbol{p}_t$ with support included in E_t and outputs the aggregated forecast

$$\widehat{y}_t = \sum_{j \in E_t} p_{j,t} f_{j,t} \,.$$

Assessment of the accuracy. We still consider in this section the quadratic loss: we define the cumulative loss and the RMSE of a strategy $\mathcal{S}$ on the first T time instances as in the previous section, that is, as, respectively,

$$\widehat{L}_T(\mathcal{S}) = \sum_{t=1}^{T} (\widehat{y}_t - y_t)^2 \qquad \text{and} \qquad \mathrm{RMSE}(\mathcal{S}) = \sqrt{\frac{1}{T} \sum_{t=1}^{T} (\widehat{y}_t - y_t)^2} \,.$$

In this section –unlike in the previous one– the evaluation of the accuracy takes place over the whole prediction period, i.e., without a training period, as the length T of this prediction period is large enough.

Comparison to the best expert or to the best constant convex combination of the experts

Things get more delicate when the corresponding quantities are defined for the experts and for convex combinations thereof. We reproduce here the methodology and definitions proposed by [FSSW97].

For a given expert. The cumulative loss of a single expert j is not a meaningful quantity but its RMSE has a natural definition:

$$\text{RMSE}(j) = \sqrt{\frac{1}{\sum_{t=1}^{T} \mathbb{I}_{\{j \in E_t\}}} \sum_{t \leqslant T : j \in E_t} (f_{j,t} - y_t)^2} \,.$$

It is also easy to introduce a notion of regret, which depends strongly on the expert j to which the strategy $\mathcal{S}$ is compared; this is why this regret is indexed by T and $\mathcal{S}$ –as before– but also by j. To get a fair comparison between j and $\mathcal{S}$ we only perform it on the time instances when j was active:

$$R_T(\mathcal{S}, j) = \sum_{t \leqslant T : j \in E_t} \left(\left(\widehat{y}_t - y_t\right)^2 - (f_{j,t} - y_t)^2 \right).$$

For a given convex combination of the experts. The final step is to extend these definitions to the case of constant convex weight vectors $\boldsymbol{q}$ so that when $\boldsymbol{q} = \delta_j$ (the Dirac mass on j) the definitions for single experts are recovered. To that end we introduce the normalization $\boldsymbol{q}^E$ of $\boldsymbol{q}$ on a subset E of $\{1, \dots, N\}$ by considering first the weight given by $\boldsymbol{q}$ to E,

$$\boldsymbol{q}(E) = \sum_{j \in E} q_j \,,$$

and second, by defining

$$\boldsymbol{q}^E = \begin{cases} (0, \dots, 0) & \text{when } \boldsymbol{q}(E) = 0; \\ \left(\dfrac{q_1 \mathbb{I}_{\{1 \in E\}}}{\boldsymbol{q}(E)}, \dots, \dfrac{q_N \mathbb{I}_{\{N \in E\}}}{\boldsymbol{q}(E)} \right) & \text{when } \boldsymbol{q}(E) > 0. \end{cases}$$

The extended definitions are then

$$\text{RMSE}(\boldsymbol{q}) = \sqrt{\frac{1}{\sum_{t=1}^{T} \boldsymbol{q}(E_t)} \sum_{t=1}^{T} \left(\sum_{j \in E_t} q_j^{E_t} f_{j,t} - y_t \right)^2 \boldsymbol{q}(E_t)}$$

and

$$R_T(\mathcal{S}, \boldsymbol{q}) = \sum_{t=1}^{T} \left(\left(\widehat{y}_t - y_t\right)^2 - \left(\sum_{j \in E_t} q_j^{E_t} f_{j,t} - y_t \right)^2 \right) \boldsymbol{q}(E_t) \,.$$

Theoretical guarantees of some aggregation strategies. [13, Section 2.3] provides an overview of the regret bounds proposed by the literature: the regrets with respect to all convex weight vectors $\boldsymbol{q}$ can be uniformly bounded by something of the order of $\sqrt{T}$, where the uniformity is over $\boldsymbol{q}$ but also over all sequences of observations y_t and experts forecasts $f_{j,t}$. The strategies providing such regret bounds are obtained by adapting the exponentially weighted average strategies based on the gradients of the losses –which where described in Section 1.2.3. The resulting strategies also rely on a parameter $\eta > 0$. For the sake of concision we do not describe them in detail and simply denote them by $\mathcal{W}_\eta^{\text{grad}}$.

Comparison to the best compound expert

The empirical studies [Gou08a, Gou08b] showed the interest of aggregation strategies tracking the best expert (via so-called compound experts) to forecast the electricity consumption.

Formal statement of the problem for specialized experts. [HW98] introduced the class of compound experts. For a given number of time instances T, this class can be identified in the case of specialized experts with the set $\mathcal{C}'_T = E_1 \times \ldots \times E_T$. For all elements $j_1^T = (j_1, \ldots, j_T)$ in $\mathcal{C}'_T$, we denote by

$$L_T(j_1^T) = \sum_{t=1}^{T} \ell(f_{j_t,t}, y_t)$$

the cumulative loss of its corresponding compound expert. Of course no strategy can achieve a cumulative loss close to the one of the best compound expert since this essentially amounts to knowing in advance the index of the best expert for the next time instance. The compound experts thus need to be constrained; we require, for instance, that they do not shift too often. Formally, the number of shifts of a given compound expert j_1^T equals

$$s(j_1^T) = \sum_{t=2}^{T} \mathbb{I}_{\{j_{t-1} \neq j_t\}}$$

and we will impose a maximal value. Shifts are also referred to as breaks in the stochastic statistical literature.

Regret with respect to compound experts with a maximal number of shifts. We denote by $\mathcal{C}'_{T,m}$ the set of all compound experts with at most m shifts. Of course, for small values of m the set $\mathcal{C}'_{T,m}$ can be empty in the context of specialized experts. It is straightforward to define the regret of a strategy $\mathcal{S}$ with respect to a compound expert j_1^T or the RMSE of a compound expert since both the strategy $\mathcal{S}$ and the expert j_1^T output a forecast at each round.

Upper bound on this regret in the case of non-specialized experts. The main reference is [HW98] and is summarized in [CBL06, Section 5.2]. It proposes an efficient sequential implementation of the strategy that essentially performs exponentially weighted averages of the cumulative losses of all compound experts, where each of the later has an initial weight that is not uniform but depends on its number of shifts. This strategy is known as the fixed-share strategy and relies on two parameters, a learning rate $\eta > 0$ and a mixing factor $\alpha \in [0, 1]$. When both parameters are well tuned (according to T and m) the regret of this strategy with respect to compound experts with at most m shifts is uniformly bounded from above by something of the order of $\sqrt{m\,T \ln N}$.

Extension to specialized experts. It is straightforward to extend the fixed-share strategy the setting of specialized experts; it can also use gradients of the losses (pseudo-losses) instead of the true losses. The details of these extensions are omitted and we simply refer the reader to [13, Section 2.3]; we underline that the extension of the fixed-share strategy to the case of specialized experts is new but absolutely natural and straightforward. Doing so we obtained two families of fixed-share strategies, $\mathcal{G}_{\eta,\alpha}$ and $\mathcal{G}^{\mathrm{grad}}_{\eta,\alpha}$, where the superscript indicates whether these strategies rely on the losses or on their gradients.

3.5.2 Presentation and characteristics of the data set

The data set considered in this section is the standard data set used for the calibration of the EDF short-term models for the French electricity load. It is described in detail in [DKO$^+$08] and [13] and we only provide an overview of its content.

It includes half-hourly electricity data and meteorological observations (temperature and cloud cover) over the whole French territory. Load data is built by EDF from the French load data measured and provided by the French national grid company, RTE ("Réseau de transport d'électricité"). Meteorological data is issued by the French weather-forecasting institution Météo-France.

Training and validation sets. This data set is divided in two parts: the first part ranges from September 1, 2002 to August 31, 2007 –we call it the training set; the second part covers the period between September 1, 2007 and August 31, 2008 –we call it the validation set. The experts we consider in this section are trained over the first part of the data set and then provide forecasts (which the strategies will aggregate) over the period corresponding to the validation set. Actually, we exclude some special days from the validation set. Out of the 366 days between September 1, 2007 and August 31, 2008, we keep 320 days. The excluded days correspond to public holidays (the day itself, as well as the days before and after it), daylight saving days and winter holidays (that is, the period between December 21, 2007 and January 4, 2008); however, we include the summer break (August 2008) in our analysis as we have access to specialized experts that are able to produce forecasts for this period. Other special days exist and correspond to temporary changes of the prices in order to reduce expected high consumption (mainly due to low temperature); they are included in the validation set whenever a preprocessing based on EDF commercial data was available. The characteristics of the consumptions y_t of the validation set are summarized in Table 3.5.

Units. The observations and forecasts of the consumption are expressed in gigawatts (GW), a unit that will generally be omitted in this section as well –in particular when we provide the values of the RMSEs.

Time intervals	Every 30 minutes
Number of days D	320
Time instances T	15 360
Number of experts N	24 $(= 15 + 8 + 1)$
Median of the y_t	56.33
Bound B on the y_t	92.76

Table 3.5. Some characteristics of the observations y_t of the French data set of operational forecasting.

Construction of three families of experts

The experts we consider here are instances of the three main categories of statistical models: parametric, semi-parametric, and non-parametric models. The reason for this choice is two-fold: first, we believe that combining base forecasters is particularly useful when they are heterogenous and exhibit significantly varied behaviors; and second, EDF could provide these three types of models. We provide below a short description of them but refer the reader to [DGS09, Section 4.1] and [13] for more details.

Parametric model. The parametric model used to generate the first group of experts is described in [BDR05] and is implemented in an EDF software called "Eventail." We mention briefly that this model is based on a nonlinear regression approach that consists of decomposing the electricity load into a main component accounting for all the seasonality of the process and a weather-dependant component. To this nonlinear regression model is added an autoregressive correction of the error of the short-term forecasts of the last seven days. Changing the parameters (the gradient of the temperature, the short-term correction) of this model, we derive 15 experts. For conciseness we refer to them as the Eventail experts.

Semi-parametric model. The second group of experts stems from a generalized additive model (henceforth referred to as the GAM model) implemented in the software `R` by the `mgcv` package developed by [Woo06]. This model is presented in [PLG09] and imports the idea of the parametric modeling presented above into a semi-parametric modeling. One of the key advantages of this model is its ability to adapt to changes in consumption habits where parametric models like Eventail need some a priori knowledge on customers behaviors. Here again, we derive different experts from the GAM model by changing the trend extrapolation effect (which accounts for the yearly economic growth) or the short-term effects like the one-day-lag effect; these changes affect the reactivity to changes along the run. Doing so, we obtain 8 experts, which we call the GAM experts.

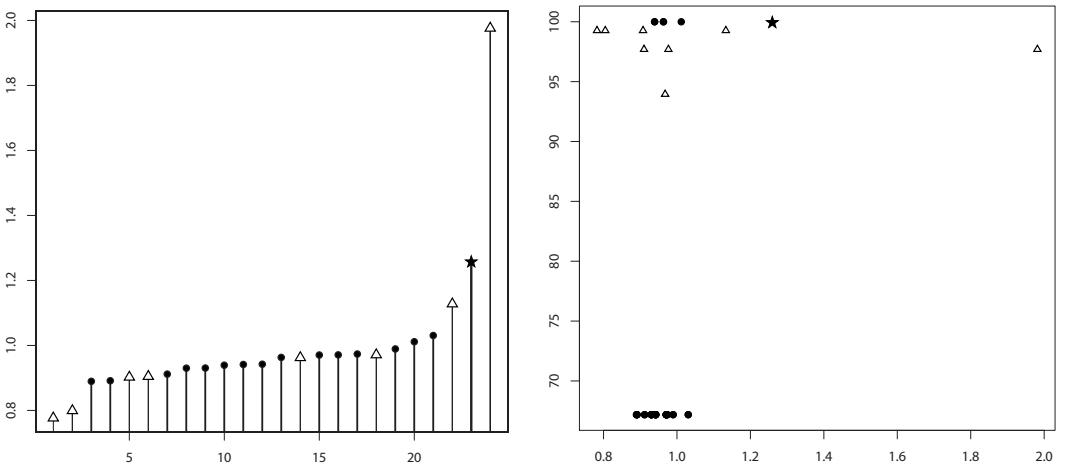

Figure 3.5. Graphical representations of the performance of the experts: sorted RMSEs (left) and RMSE–frequency-of-activity pairs (right); Eventail experts are depicted by the symbols •, GAM experts are represented by △ while ⋆ stands for the similarity expert.

Non-parametric model. The last expert is drastically different from the two previous groups of experts and its construction is presented in [APS06] and [ABCP10]. It relies on a univariate method (i.e., it does not require any exogenous factor like weather conditions); the key idea is to assume that the load is driven by an underlying stochastic curve and to model each day as a discrete recording of this functional process at half-hourly instances. Forecasts are then performed according to a similarity measure between days. We call this expert the similarity expert.

Performance of the experts. The characteristics of the experts presented above are depicted in Figure 3.5. The bar plot represents the (sorted) values of the RMSEs of the 24 available experts. The scatter plot relates the RMSE of each of the expert to its frequency of activity, that is, it plots the pairs

$$\left(\text{RMSE}(j),\ \frac{\sum_{t=1}^{T} \mathbb{I}_{\{j \in E_t\}}}{T} \right)$$

for all experts j.

Out of the 15 Eventail experts, 3 are permanently active; they correspond to the operational model used by EDF and to two variants of it based on different short-term corrections. The other 12 Eventail experts are inactive during the summer as their forecasts are redundant with the operational model (they were obtained by changing the value of the gradient of the temperature, which affects the winter forecasts only). GAM expert are active on an overwhelming fraction of the time and are sleeping only during periods when R&D practitioners know beforehand that they will perform poorly (e.g., in time periods close to public holidays); the lengths of these periods depend on

Name of the reference strategy	Formula	Value
Uniform sequential aggregation rule	$\textsc{rmse}(\mathcal{U})$	$= 0.724$
Uniform convex weight vector	$\textsc{rmse}\big((1/24, \ldots, 1/24)\big)$	$= 0.748$
Best single expert	$\min_{j=1,\ldots,24} \textsc{rmse}(j)$	$= 0.782$
Best convex weight vector	$\min_{\boldsymbol{q}\in\mathcal{P}} \textsc{rmse}(\boldsymbol{q})$	$= 0.683$
Best compound expert		
Size at most $m = 50$	$\min_{j_1^T \in \mathcal{C}'_{T,50}} \textsc{rmse}\big(j_1^T\big)$	$= 0.534$
Size at most $m = 100$	$\min_{j_1^T \in \mathcal{C}'_{T,100}} \textsc{rmse}\big(j_1^T\big)$	$= 0.474$
Prescient strategy	$\min_{j_1^T \in E_1 \times E_2 \times \ldots \times E_T} \textsc{rmse}\big(j_1^T\big)$	$= 0.223$

Table 3.6. Definition and performance of several reference strategies on the data set of electricity consumption.

the parameters of the expert, hence the frequencies of activity vary among the GAM experts. Finally, the similarity expert is always active.

An operational constraint

The operational constraint consists of outputting half-hourly forecasts every day at 12:00 for the next 24 hours –that is, of forecasting simultaneously the 48 next time instances. All models presented above are assumed to abide by this constraint so that the aggregation strategies may also do so. However, unlike in the previous empirical study, we do not impose that the latter use the same weight vector to combine the experts forecasts on these time instances. Put differently, the experts forecasts can be aggregated with weight vectors that depend on the moment of the day; it is even necessary to do so when some experts get inactive or active during the set of time instances for which predictions are required.

3.5.3 Performance of some aggregation strategies

Performance of some reference strategies

Table 3.6 shows that the constructed experts exhibit an excellent performance in view of the typical orders of magnitude of the y_t indicated in Table 3.5. Some reference strategies deserve comments.

Uniform sequential aggregation versus use of the uniform convex weight vector. In the setting of specialized experts there is a subtle difference between the uniform sequential aggregation strategy $\mathcal{U}$ and the use of the uniform convex weight vector $\boldsymbol{q} = (1/24, \ldots, 1/24)$. The former chooses indeed at each time instance t the convex weight vector given by the uniform distribution over the set E_t of active experts, so that

$$\text{RMSE}(\mathcal{U}) = \sqrt{\frac{1}{T}\sum_{t=1}^{T}\left(\frac{\sum_{j\in E_t} f_{j,t}}{|E_t|} - y_t\right)^2}$$

$$\text{while} \qquad \text{RMSE}\big((1/24, \ldots, 1/24)\big) = \sqrt{\frac{1}{\sum_{t=1}^{T}|E_t|}\sum_{t=1}^{n}|E_t|\left(\frac{\sum_{j\in E_t} f_{j,t}}{|E_t|} - y_t\right)^2}.$$

Thus, in the assessment of the performance of the uniform convex weight vector the losses associated with time instances for which many experts are active count more than those for which few experts only are active; for the uniform sequential aggregation strategy $\mathcal{U}$ all losses have the same weight.

Reference values. The use of the uniform convex weight vector leads to a RMSE larger than the one of the strategy $\mathcal{U}$, which means that experts tend to be more active at time instances when the forecasting is more difficult. This is an advantage from which aggregation strategies will profit. But for the time being the consequence is that the RMSE of the best expert is larger than the one of the most naive aggregation strategy, $\mathcal{U}$. Table 3.6 thus indicates that our more sophisticated aggregation strategies should improve significantly on the strategy $\mathcal{U}$ (whose RMSE is 0.724). The performance of the best constant convex weight vector is already slightly better (its RMSE equals 0.683) but the RMSEs achieved by the compound experts show that strong improvements upon $\mathcal{U}$ are possible.

The end of this section illustrates that such an improvement takes place for the exponentially weighted average strategy $\mathcal{W}_\eta^{\text{grad}}$ using the gradients of the losses, as well as for the fixed-share strategies $\mathcal{G}_{\eta,\alpha}$ and $\mathcal{G}_{\eta,\alpha}^{\text{grad}}$, which were briefly mentioned at the end of Section 3.5.1.

Performance and robustness properties of the studied aggregation strategies

To tabulate the performance of the family of strategies $\mathcal{W}_\eta^{\text{grad}}$ we resorted to a grid of 19 parameters η logarithmically evenly spaced between 10^{-6} and 1. For the families $\mathcal{G}_{\eta,\alpha}$ and $\mathcal{G}_{\eta,\alpha}^{\text{grad}}$ we chose a finite grid in $\mathbb{R}_+ \times [0,1]$ containing 22×6 points. We summarize in Table 3.7 the obtained performance; for each family we only report the RMSEs corresponding either to the best constant choices of a point in the grids or to the tuned meta-strategies using these grids and based on the three families.

Comments. The three families of strategies obtain satisfactory –and even quite good– results compared to the reference values indicated in the comments relative to Table 3.6.

		Best constant parameter(s)	Tuned on the grid
RMSE of	$\mathcal{W}_{\eta}^{\mathrm{grad}}$	0.650	0.654
	$\mathcal{G}_{\eta,\alpha}$	0.632	0.644
	$\mathcal{G}_{\eta,\alpha}^{\mathrm{grad}}$	0.598	0.599

Table 3.7. RMSEs of three aggregation strategies on the data set of French electricity consumption: with the best constant parameters (left column) and when tuned on the grids described in Section 3.5.3 (right column).

Here again gradient-based strategies are more efficient than their counterparts using directly the losses –which was expected in view of the theoretical results presented in Section 1.2.3. We actually were pleasantly surprised by the performance of the family $\mathcal{G}_{\eta,\alpha}^{\mathrm{grad}}$ and even found them intriguing as the following robustness remarks underline.

Robustness study. This study –which we simply mentioned in the case of the forecasting of ozone peaks in Section 3.4.3– consists of locally comparing the performance of the aggregation strategies to the one of the best expert or of the best constant convex weight vector. The RMSE is indeed a global criterion and we want to check that the overall good performance does not come at the cost of local disasters in the accuracy of the aggregated forecasts. To that end we split the data set by the half-hours into 48 sub-data sets; for each of these subsets we compute the RMSEs of the strategies discussed above and study also the scattering of the absolute values of the prediction residuals.

The latter are defined as $|\widehat{y}_t - y_t|$, where y_t denotes the observed consumption at time instance t and $\widehat{y}_t$ is its aggregated forecast. We focus on the quantiles of these prediction residuals –and more particularly on the ones of orders 75 % or 90 %, whose values measure the extent of disastrous forecasts. Figure 3.6 depicts the performance of the tuned meta-strategies based respectively on the families $\mathcal{W}_{\eta}^{\mathrm{grad}}$ and $\mathcal{G}_{\eta,\alpha}^{\mathrm{grad}}$. The first meta-strategy has uniformly similar or slightly smaller RMSEs and quantiles than the ones of the best constant convex weight vector. The performance of the second meta-strategy is intriguing: its accuracy is significantly improved with respect to the one of the best constant convex weight vector between 12:00 and 21:00 but is also slightly worse than the latter between 6:00 and 12:00. We can provide no reason for this behavior yet; the aim would be to take advantage of the improvements that arise right after the update performed at noon, maybe by checking whether another update at midnight would be useful (though it would not satisfy the aforementioned operational constraint).

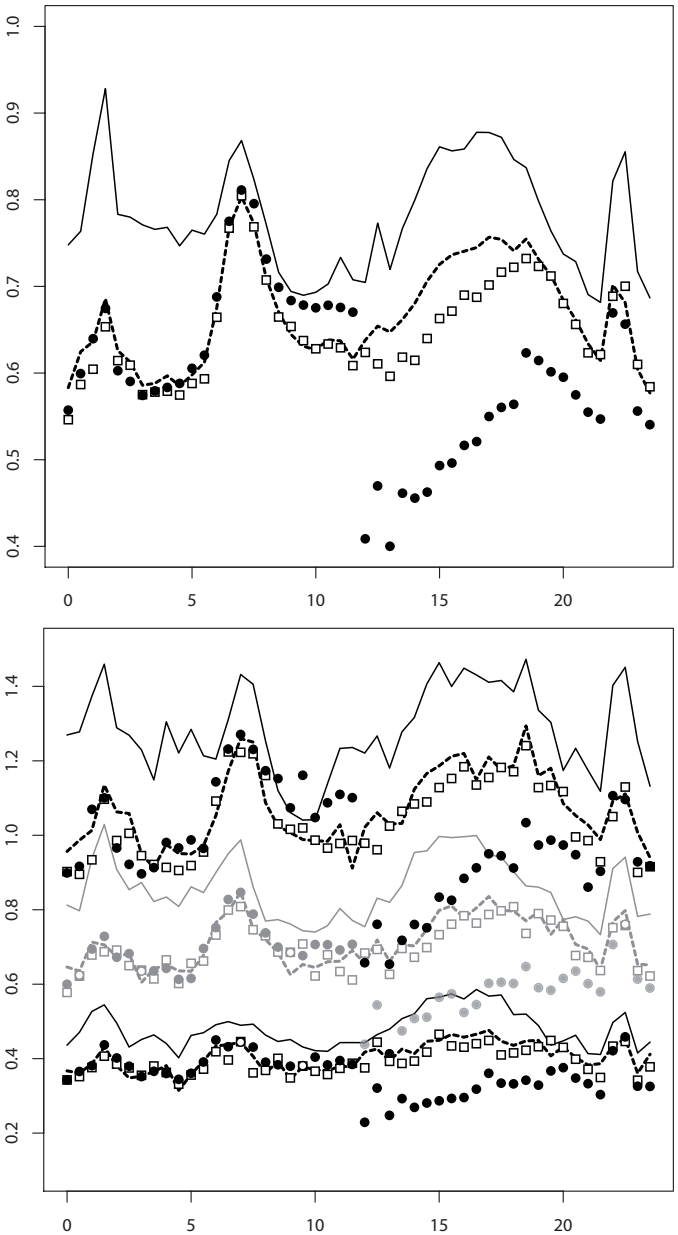

Figure 3.6. Measures of the half-hourly performance of the overall best expert (solid line) and of the overall best convex weight vector (dashed line), as well as of the tuned meta-strategies based on the families $\mathcal{W}_{\eta}^{\mathrm{grad}}$ (symbol: •) and $\mathcal{G}_{\eta,\alpha}^{\mathrm{grad}}$ (symbol: □); RMSEs (top picture) and quantiles of orders 50 % (black), 75 % (grey), and 90 % (black) of the absolute values of the residuals (bottom picture). Axes: x–axes index the half hours; y–axes measure respectively the RMSEs (top picture) and the absolute values of the residuals (bottom picture).

3.6 Conclusions et research perspectives

3.6.1 Conclusions for practitioners

This chapter was devoted to the (good) operational performance of some strategies that sequentially aggregate experts forecasts. We described a general methodology as well as some new aggregation strategies and variants of existing strategies: windowing, use of discounted losses, automatic sequential tuning of the parameters on grids. We then instantiated these results –with some minor adaptations– for two applications: the forecasting of daily ozone peaks and of half-hourly electricity consumption. However, we encourage the reader to use the strategies described in this chapter and in Chapter 1 in all settings of sequential forecasting where he would have at his disposal a set of experts within which he cannot guess in advance the best member. These experts can in particular be given by methods relying on some stochastic modeling and parameterized by several parameters: an alternative to the tuning of all these parameters is to consider several instances of the method tuned each with a different enough set of parameters. Another option is to consider on top of such experts other experts for which no theoretical guarantee exists but which are supported by some intuition. The two empirical studies illustrated the proverb "Garbage in, garbage out": whenever there exist some good experts the aggregation strategies also exhibit a good performance. In particular we do not need to construct only good experts and simply have to ensure that some of the experts (which do not need to be known in advance) will output accurate forecasts.

3.6.2 Research perspectives

At the methodological level

We underlined in Section 3.1 that for two procedures with good empirical performance no theoretical bound was established yet: the tuned meta-strategies of Section 3.1.1 and the sequential Lasso forecaster of Section 3.1.3.

For the forecasting of ozone peaks

The short-term projects are first to evaluate the impact of the pre-treatment on the experts to reduce their biases and second to study how the performance of the aggregation strategies evolves when the prediction period gets longer (one-year long) with the same number of experts or more of them. Preliminary results showed that –as expected– the gains in performance with respect to the best expert or to the best constant convex weight vector are even more significant in this case.

A mid-term project is to study a sequential classification problem. We recall that French authorities need respectively to inform and to alert people whenever the concentrations are expected to be above 180 $\mu g\,m^{-3}$ and 240 $\mu g\,m^{-3}$; we therefore need to sequentially forecast the class (low concentration, information required, alert required) to which the next peak will belong. Here also we ran some preliminary experiments

but were unable to perform better than the procedure that consists of comparing the aggregated forecasts to the thresholds 180 or 240. This is quite surprising since classification problems are usually easier to handle than providing accurate forecasts of quantitative values.

We also mention a recent study by Vivien Mallet on the links between data assimilation and aggregation of experts [Mal10].

For the forecasting of electricity consumption

The short-term aim is to better understand why the fixed-share strategies are so accurate in the short run but achieve more disappointing results at the end of the prediction round. We should also take a better advantage of the specialization of the experts by constructing them more carefully and in a more systematic way (e.g., having more experts dedicated to winter or to summer). Finally, at a methodological level we would like to extend the regularized sequential linear regression forecasters of Section 3.1.3 to the setting of specialized experts. To the best of our knowledge this extension was not considered yet and seems uneasy; some preliminary attempts described in [DGS09] failed (as far as the empirical performance on the data set is considered: we expect significant improvements from such strategies compared to strategies constrained to use convex weight vectors).

In addition it would be interesting –here but also for the forecasting of ozone peaks– to be able to measure the uncertainties on the aggregated forecasts. Such measures could rely either on the scattering of the experts forecasts themselves (the uncertainties are larger as this scattering is larger) or on measures of uncertainties on their forecasts provided directly by the experts.

Other data sets

Some other data sets could be studied, e.g., the forecasting of exchange rates in economics –where experts would be given by the predictions of several financial analysts– or of river heights in hydrogeology. My current work in these two fields is to identify reliable and competent partners in each field who will construct the experts and evaluate the results achieved by the aggregation strategies. The longer-term aim would then be to provide a software implementing the strategies discussed in this chapter, once the latter have been tested on varied enough data sets and all have fully automatic variants.

CHAPTER 4

Stochastic continuum-armed bandit problems and miscellaneous contributions

INTRODUCTION. This final chapter is devoted to contributions that are not linked to the main focus of this thesis, namely, the sequential prediction of arbitrary sequences. Two such contributions lie within a research field which I recently started studying –the stochastic continuum-armed bandit problems.

Table of contents

4.1 Stochastic continuum-armed bandit problems [8, 11]

We describe first the most general model where the arms are indexed by an arbitrary (finite or infinite) set $\mathcal{X}$; with each arm is associated a probability distribution. We then state some of the obtained results, essentially those for the case where $\mathcal{X}$ is a metric space. (We thus omit the first half of [8], which is devoted to the case of a finite set $\mathcal{X}$.)

Stochastic setting. As will become clear with the review of the literature proposed below, there exists a version of the multi-armed bandit problem with finitely (or even countably) many arms in the setting of randomized prediction of arbitrary sequences discussed in Section 1.1.3; forecasters performing well in this setting can be constructed based on techniques similar to the ones presented in Section 1.3.1. But for once, the rest of this section is devoted to a good old stochastic setting.

4.1.1 Mathematical description of the model (bounded payoffs)

A set of arms indexed by $\mathcal{X}$ is available and a statistician plays against a stochastic environment E according to the protocol described in Figure 4.1. In particular he gets

Parameters: a known set of arms $\mathcal{X}$ (finite or infinite; equipped with a topology); an unknown environment $E : \mathcal{X} \to \Delta\big([0,1]\big)$

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

1. The statistician chooses a probability distribution $\nu_t \in \Delta(\mathcal{X})$ and draws an arm $I_t \in \mathcal{X}$ at random according to ν_t;
2. The environment draws the statistician's payoff Y_t independently at random according to the distribution $E(I_t)$ associated with the chosen arm;
3. The statistician only observes Y_t (and recalls this quantity for the next rounds).

Figure 4.1. The protocol of the multi-armed bandit problem with arms indexed by $\mathcal{X}$.

at each round a bounded payoff with possible values in a range known beforehand; for simplicity we assume that this range is the interval $[0,1]$. We denote by $\Delta([0,1])$ the set of all probability distributions over $[0,1]$.

Notion of stochastic environment. A stochastic environment E is defined as a mapping $\mathcal{X} \to \Delta([0,1])$. It associates with each element $x \in \mathcal{X}$ a probability distribution $E(x)$ over $[0,1]$. When the statistician chooses the arm $I_t \in \mathcal{X}$ at time instance t, the environment draws a payoff Y_t independently at random according to the distribution $E(I_t)$. We denote by $\mu_E : \mathcal{X} \to [0,1]$ the mean-payoff function, which associates with each $x \in \mathcal{X}$ the expectation $\mu_E(x)$ of the probability distribution $E(x)$.

Definition of a strategy of the statistician. The statistician ignores which environment E he is playing with. The only information at his disposal are given by the payoffs associated with the arms chosen in the past. He thus determines the arm I_t to pull at a given time instance $t \geqslant 2$ depending on the arms $I_1, \ldots, I_{t-1}$ and their associated payoffs $Y_1, \ldots, Y_{t-1}$; he does so possibly at random thanks to an auxiliary randomization. To that end we assume that $\mathcal{X}$ is a topological space, equipped with its Borel σ–algebra, and denote by $\Delta(\mathcal{X})$ the set of probability distributions over $\mathcal{X}$.

A strategy Ψ is thus given by some initial distribution $\Psi_1 \in \Delta(\mathcal{X})$ and a sequence of measurable mappings Ψ_t, where $t \geqslant 2$. For each $t \geqslant 2$ the mapping Ψ_t is defined over $\mathcal{X}^{t-1} \times [0,1]^{t-1}$ and takes its values in $\Delta(\mathcal{X})$: with the notation of Figure 4.1, the statistician then chooses the probability distribution

$$\nu_t = \Psi_t\Big(I_1, \ldots, I_{t-1}, Y_1, \ldots, Y_{t-1}\Big)$$

over $\mathcal{X}$ and draws his arm I_t at random according to ν_t.

Auxiliary randomizations. The statistician and the environment both resort to sequences of auxiliary randomizations. The probabilities $\mathbb{P}$ and expectations $\mathbb{E}$ will all be relative to these randomizations only (they depend neither on E nor on Ψ).

Aim: statement and difficulties in achieving it

The aim of the statistician is to ensure that his cumulative payoff $Y_1 + \ldots + Y_T$ is as large as possible.

Dilemma between exploration and exploitation. To do so and because the feedback obtained when choosing a given arm is random he must perform a trade-off between exploration and exploitation. Exploration means pulling each arm a significant number of times in order to estimate accurately its associated probability distribution; exploitation means using the gained information to pull more frequently the better arms. Since both need to be performed simultaneously a dilemma between exploration and exploitation occurs.

Notion of regret and reformulation of the aim: minimization of the regret

Here again, ensuring that some regret is small entails a large cumulative payoff. Actually, in the considered stochastic setting all theoretical results are formulated in the literature in terms of expected cumulative payoffs, so that the regret of a strategy itself will be studied as a deterministic quantity (corresponding to some expectation).

Formally, the regret of a strategy Ψ against an environment E is equal to

$$R_T(\Psi, E) = T\mu_E^\star - \sum_{t=1}^{T} Y_t \qquad \text{where} \quad \mu_E^\star = \sup_{x \in \mathcal{X}} \mu_E(x)\,;$$

but only the expected quantities $\mathbb{E}\big[R_T(\Psi, E)\big]$ will be bounded. Minimizing the regret is indeed equivalent to maximizing the expected cumulative payoff.

Regret and pseudo-regret. By the tower rule,

$$\mathbb{E}\left[\sum_{t=1}^{T} Y_t\right] = \mathbb{E}\left[\sum_{t=1}^{T} \mu_E(I_t)\right] \leqslant T \sup_{x \in \mathcal{X}} \mu_E(x) = T\mu_E^\star\,.$$

Therefore, it always holds that $\mathbb{E}\big[R_T(\Psi, E)\big] \geqslant 0$: the regret has a clear interpretation in this chapter, it can only be nonnegative. We then define the pseudo-regret of a strategy Ψ against an environment E as the unobserved quantity

$$R'_T(\Psi, E) = \sum_{t=1}^{T} \big(\mu_E^\star - \mu_E(I_t)\big)\,;$$

the equality $\mathbb{E}\big[R_T(\Psi, E)\big] = \mathbb{E}\big[R'_T(\Psi, E)\big]$ is entailed again by the tower rule. In the proofs of the theoretical bounds on the expected regret it is often more convenient to consider the pseudo-regret instead of the regret.

Why $T\mu_E^\star$ is not replaced by a certain supremum of empirical processes. In the considered setting we do not impose that at each round a realization of each of the distributions $E(x)$ be drawn, for all $x \in \mathcal{X}$; only a realization associated with each chosen is drawn. It is wise to do so, for measurability issues would arise otherwise since the set of arms $\mathcal{X}$ is arbitrary and may be non countable. However, this prevents us from replacing each of the $T\mu_E(x)$ by a sum of random variables indexed by x and hence from replacing $T\mu_E^\star$ by a supremum of empirical processes each indexed by x. In conclusion, we consider $T\mu_E^\star$ for want of anything better.

Reformulation of the aim in terms of the regret. We assume that the statistician has some knowledge on the environment E he is playing against: he knows that E belongs to some (possibly non-parametric) family $\mathcal{F}$ of mappings $\mathcal{X} \to \Delta\big([0,1]\big)$. Like in the previous chapters we aim at ensuring that the (expected) per-round regret is asymptotically non positive; since we showed above that it is nonnegative in the present setting, the aim is therefore that this expected per-round regret tends to zero.

We thus say that a strategy Ψ minimizes its regret with respect to a family $\mathcal{F}$ if

$$\forall\, E \in \mathcal{F}, \qquad \lim_{T\to\infty} \frac{\mathbb{E}\big[R_T(\Psi, E)\big]}{T} = 0\,.$$

In some cases the convergences above towards 0 hold uniformly over $\mathcal{F}$.

(Extremely brief) literature review

The problem was first mentioned by Robbins [Rob52]. We can split the literature in two main categories: the articles dealing with the case where $\mathcal{X}$ is a finite or a countable set (with two associated sub-categories, depending on whether the payoffs are given by random variables or by arbitrary sequences) and the ones about uncountably many arms.

Bandit problems with finitely (or countably) many arms. In the finite case one usually denotes by $|\mathcal{X}| = K$ the cardinality of $\mathcal{X}$. The most important results for the stochastic case presented above are that the regret can indeed be minimized against all environments E, that is, against all K–tuples of probability distributions over $[0,1]$, with the following rates of convergence towards 0: a constant depending on E times $(\ln T)/T$ for the strategies proposed in [LR85, BK96, ACBF02, AB09, HT10] and uniform rates of the order of $\sqrt{K(\ln T)/T}$ for [ACBF02], $\sqrt{K(\ln K)/T}$ for [ACBFS02], and $\sqrt{K/T}$ for [AB09]. The latter uniform rate is the optimal uniform rate as follows from the lower bound stated in [ACBFS02]. The above convergence results can be extended thanks to the doubling trick to cover the cases where countably many arms are available –at the cost of loosing the uniformity with respect to the environments E in the bounds.

There also exists a version of the problem for arbitrary sequences, whose protocol is close to the one stated in Section 1.1.3; this more difficult setting is tackled, e.g., by [ACBFS02, AB09].

Bandit problems with uncountably many arms. This setting was first considered by [Agr95, Kle04] and later studied by [Cop09, AOS07, KSU08]. The strategies proposed in these articles only minimize the regret under some topological assumptions over $\mathcal{X}$ and against classes of smooth enough environments E (where the smoothness assumptions are actually on μ_E).

For instance, [KSU08] assumes that $\mathcal{X}$ is a metric space and exhibits strategies minimizing the regret against the class of environments E with mean-payoff functions μ_E that are L–Lipschitz, with a Lipschitz constant L smaller than a given bound L_0 known by the statistician. That is, assumptions on the global smoothness of the functions μ_E are issued.

In contrast, the assumptions in [AOS07] only refer to the local behavior of μ_E around its global maxima.

4.1.2 Necessary and sufficient conditions for the minimization of the regret [8]

The second half of [8] considers the following families of environments:

$$\mathcal{F}_{\text{all}} = \Delta\big([0,1]\big)^{\mathcal{X}} \qquad \text{and} \qquad \mathcal{F}_{\text{cont}} = \mathcal{C}\Big(\Delta\big([0,1]\big)^{\mathcal{X}}\Big)\,,$$

which are respectively the sets of all possible environments and of the environments E whose associated mean-payoff functions μ_E are continuous. We characterize the existence of strategies minimizing the regret against these families.

Theorem 4.1. *When $\mathcal{X}$ is a metric space, the regret can be minimized against all environments of the family $\mathcal{F}_{\text{cont}}$ if and only if $\mathcal{X}$ is separable.*

Proof sketch. The proof only requires writing the following ideas in a formal way; the ideas are all relative to the possibility or the impossibility of a uniform exploration over $\mathcal{X}$.

On the one hand, when $\mathcal{X}$ is separable there exists by definition a dense countable subset of the arms and it suffices to pull only arms in the latter as their performance is close (by continuity and density) to the one of the best arms in $\mathcal{X}$. But we already mentioned that the regret can be minimized against all environments whenever countably many arms only are available. Indeed, to do so, one simple (and somewhat suboptimal) way is to proceed in successive regimes, by alternating between exploration phases – resorting to a probability distribution giving a positive probability to each arm in the mentioned dense countable subset– and exploitation phases.

Reciprocally, a non-separable metric space contains uncountably many disjoint balls $\mathcal{B}(a,\rho)$ for some common radius $\rho > 0$ and with centers given by $a \in A$ (the set A being uncountable). We associate each of these balls with an environment E_a whose mean-payoff function μ_{E_a} has a maximum equal to 1 and a support included in $\mathcal{B}(a,\rho)$. Now, each probability distribution over $\mathcal{X}$ can put a positive probability on at most countably many such balls. Thus, a given strategy will obtain null payoffs at each round against all environments E_a, with $a \in A$, except maybe against a countable number of

them (the ones whose supports it had a positive probability to explore). Its regret will thus be equal, for most of the environments, to the number of rounds and hence will not be sublinear.

Corollary. We adopt an "à la Bourbaki" approach and derive the following result from Theorem 4.1.

Corollary 4.2. Let $\mathcal{X}$ be an arbitrary set. The regret can be minimized against the family $\mathcal{F}_{\text{all}}$ of all possible environments if and only if $\mathcal{X}$ is countable.

We already indicated above that whenever $\mathcal{X}$ is countable, the regret can be minimized against all environments. We only need to show that this condition is necessary. Actually, the corollary above is simply an instantiation of Theorem 4.1: when $\mathcal{X}$ is equipped with the discrete topology (which corresponds to the Hamming distance) all applications $\mathcal{X} \to [0,1]$ are continuous, so that $\mathcal{F}_{\text{cont}} = \mathcal{F}_{\text{all}}$.

Conclusion. Minimal topological assumptions on $\mathcal{X}$ and/or smoothness assumptions on the mean-payoff functions μ_E are necessary to ensure the existence of strategies minimizing the regret. To exhibit simple and efficient such strategies we will however need to strengthen somewhat these assumptions.

4.1.3 An efficient hierarchical strategy minimizing the regret [11]

We introduced a strategy called HOO (which stands for "hierarchical optimistic optimization") and relying on three parameters.

Parameters of HOO. This strategy relies on two real numbers $\nu_1 > 0$ and $\rho \in]0,1[$, as well as on a tree of coverings $\mathcal{T} = (\mathcal{T}_{h,i})$, that is, on a collection of (non necessarily disjoint) subsets of $\mathcal{X}$ indexed by $h \in \mathbb{N}$ and $1 \leqslant i \leqslant 2^h$ and satisfying

$$\begin{aligned} &\mathcal{T}_{0,1} = \mathcal{X}\,, \\ &\mathcal{T}_{h,i} = \mathcal{T}_{h+1,2i-1} \cup \mathcal{T}_{h+1,2i} \qquad \text{for all } h \geqslant 0 \text{ and } 1 \leqslant i \leqslant 2^h. \end{aligned}$$

For all depths $h \geqslant 0$, the subsets $\mathcal{T}_{h,i}$ cover $\mathcal{X}$ when i varies in $\{1, \dots, 2^h\}$ –hence the name of tree of coverings for $\mathcal{T}$.

High-level principle. We only describe HOO in an informal manner. With each node (h,i) in $\mathcal{T}$ we associate an estimator of the supremum of μ_E on the subset $\mathcal{T}_{h,i}$. This estimator is defined recursively based on the estimators associated with the children nodes $(h+1, 2i-1)$ and $(h+1, 2i)$ of (h,i). At each round t the strategy chooses and expands the most promising path: it starts from the root and chooses at each node the child whose associated estimator has the largest value. When it attains a node (H_t, J_t) that was never explored (and with which no estimator is associated yet) it pulls an arm I_t at random in the subset $\mathcal{T}_{H_t,J_t}$; after getting its payoff it can then construct an

estimator for this node –even if it will be rather crude for the time being since it is based on one observation only.

References. This hierarchical strategy relying on a tree is inspired by the techniques and algorithms presented in [KS06, GWMT06, CM07].

Dissimilarity function and assumptions on the parameters of HOO. A dissimilarity function ℓ is a mapping $\mathcal{X}^2 \to \mathbb{R}_+$ such that $\ell(x,x) = 0$ for all $x \in \mathcal{X}$ (but that is neither necessarily symmetric, nor does necessarily satisfy the separation axiom or the triangle inequality). We denote by $\mathcal{B}(x,r)$ the ball with center x and ℓ–radius r and consider the following set of assumptions.

Assumption 4.3. The parameters of HOO are chosen so that there exists a dissimilarity function ℓ and a real number $\nu_2 > 0$ such that, for all integers $h \geqslant 0$,

(a) for all $1 \leqslant i \leqslant 2^h$, the diameter of $\mathcal{T}_{h,i}$ satisfies $\sup_{x,y \in \mathcal{T}_{h,i}} \ell(x,y) \leqslant \nu_1 \rho^h$;

(b) for all $1 \leqslant i \leqslant 2^h$, there exists $x_{h,i} \in \mathcal{T}_{h,i}$ such that $\mathcal{B}_{h,i} \stackrel{\text{def}}{=} \mathcal{B}(x_{h,i}, \nu_2 \rho^h) \subseteq \mathcal{T}_{h,i}$;

(c) for all $1 \leqslant i < j \leqslant 2^h$, the balls $\mathcal{B}_{h,i}$ and $\mathcal{B}_{h,j}$ are disjoint.

Environments with $(1,\ell)$–weakly Lipschitz mean-payoff functions. For all dissimilarity functions ℓ, we denote by $\mathcal{F}_{1,\ell}$ the class of environments E with mean-payoff functions μ_E satisfying

$$\forall\, (x,y) \in \mathcal{X}^2, \qquad \mu_E^\star - \mu_E(y) \leqslant \mu_E^\star - \mu_E(x) + \max\Big\{\mu_E^\star - \mu_E(x),\ \ell(x,y)\Big\}$$
$$\text{where} \quad \mu_E^\star = \sup_{x \in \mathcal{X}} \mu_E(x)\,.$$

Such functions μ_E are said weakly Lipschitz with respect to ℓ, with weak Lipschitz constant equal to 1. Indeed, whenever μ_E is Lipschitz (in the classical sense) with respect to ℓ with Lipschitz constant equal to 1, it is in particular weakly Lipschitz.

An example of an obtained uniform regret bound. [11] is essentially devoted to improving the orders of magnitude of the regret bounds under certains conditions on $\mathcal{X}$ and μ_E that are weaker than the ones considered in [KSU08]. For the sake of simplicity we only report below one special case of our obtained bounds; a strong enough topological assumption on $\mathcal{X}$ and some smoothness assumptions on μ_E entail the existence of a uniform regret bound with respect to quite large a class of environments. Here again, the assumptions considered for this uniform bound are somewhat weaker than the ones of [KSU08], which required, for instance, that the mean-payoff functions μ_E be Lipschitz (in the classical sense) with respect to a metric d over $\mathcal{X}$. The topological assumption is in terms of the ℓ–packing dimension of $\mathcal{X}$.

Definition 4.4. The ε–packing number $\mathcal{N}(\mathcal{X}, \ell, \varepsilon)$ of $\mathcal{X}$ with respect to the dissimilarity function ℓ is the size of the largest packing of $\mathcal{X}$ with disjoint ℓ–open balls of radius ε. The ℓ–packing dimension of $\mathcal{X}$ is then defined as

$$D_{\mathcal{X},\ell} = \limsup_{\varepsilon \to 0} \frac{\ln \mathcal{N}(\mathcal{X}, \ell, \varepsilon)}{\ln(1/\varepsilon)} \,.$$

Theorem 4.5. *We consider the strategy* HOO *tuned with some fixed parameters ν_1, ρ, and $\mathcal{T}$. Then, for all dissimilarity functions ℓ satisfying Assumption 4.3 and for all real numbers $D > D_{\mathcal{X},\ell}$,*

$$\limsup_{T \to \infty} \frac{\sup_{E \in \mathcal{F}_{1,\ell}} \mathbb{E}\big[R_T(\text{HOO}, E)\big]}{T^{(D+1)/(D+2)} (\ln T)^{1/(D+2)}} < \infty \,.$$

Optimality of this bound. We then show in [11] that when the dissimilarity function ℓ is a distance, the order of magnitude in T of the obtained uniform bound cannot be improved in general. To do so we explain how the regret of a strategy designed for the $\mathcal{X}$–armed case can be interpreted as the regret of some induced strategy in the setting of finitely many armed (K–armed) bandit problems; and then resort to the lower bound result proved in [ACBFS02]. We note that [KSU08] also offers a similar optimality result, based however on somewhat different proof techniques.

4.1.4 Perspectives for future research: adaptation to the unknown smoothness parameters

For the time being the literature on stochastic continuum-armed bandit problems is at a somewhat preliminary stage. Theorem 4.5 is typical of the results exhibited so far therein: a strategy –parameterized by certain fixed parameters– is considered and it is shown that it minimizes its regret at least with respect to some large class of environments. However, the latter class –though often massive– is typically defined in terms of the parameters of the strategy and/or of some additional unknown parameters.

This is why we explicitly indexed the classes of environments by these parameters: the dissimilarity function ℓ and the weak Lipschitz constant, equal to 1. In particular the HOO strategy exploits the knowledge that this constant equals 1 and does not aim at estimating it (or, more generally, at estimating the smoothness of the underlying mean-payoff functions).

Changing the viewpoint. A statistician would in contrast consider that the environment E comes first and that the strategies should be constructed adaptively to it. However, a statistical model could be available, in which case the statistician would know beforehand that the environments have smooth enough mean-payoff functions μ_E, e.g., weakly Lipschitz with respect to some dissimilarity function ℓ and with some Lipschitz constant L. These parameters ℓ and L are unknown and the aim is therefore to sequentially tune the parameters of the strategies so that their regrets are minimized despite all.

Ideally uniform regret bounds similar to the one of Theorem 4.5 could be proved for these adaptive strategies; the differences in the orders of magnitude would then measure the price for the adaptation. An even more difficult issue would be to ignore the payoff range (here, we assumed throughout the chapter that it was $[0, 1]$).

Connections between machine learning and adaptive statistics. This change of viewpoint and the adaptation to the unknown smoothness parameters of the environment are an opportunity to connect the setting and the techniques considered in machine learning with the ones of adaptive statistics. A preliminary idea in this respect was proposed by Pascal Massart: approximation theory states how well smooth functions can be represented by histograms; now, a long series of results in classical statistics indicates how to estimate histograms, e.g., with model-selection based procedures; we would therefore have first to transpose these estimation results of the classical setting into the framework of stochastic bandits.

4.2 Miscellaneous works [14, 15]

I only briefly describe their respective focuses. [14] is an empirical economic study dedicated to the assessment of the quality of some macro-economic data; it essentially consists of the application of many χ^2 goodness-of-fit tests, as well as a thorough review of the mathematical literature on Benford's law. [15] is a textbook aimed at graduate students (in French, written with Vincent Rivoirard); it presents a modern and concise view on statistics, with applications to some problems in machine learning (classification, data compression, two-armed bandit problems) and to some other problems in statistics (non parametric estimation of regression functions or of density functions, censored data).

Bibliography

[AB09] J.-Y. Audibert and S. Bubeck. Minimax policies for adversarial and stochastic bandits. In *Proceedings of the Twenty-Second Annual Conference on Learning Theory (COLT)*, 2009.

[ABCP10] A. Antoniadis, X. Brossat, J. Cugliari, and J.M. Poggi. Clustering functional data using wavelets. In *Proceedings of the Nineteenth International Conference on Computational Statistics (COMPSTAT)*, 2010.

[ABR07] J. Abernethy, P.L. Bartlett, and A. Rakhlin. Multitask learning with expert advice. In *Proceedings of the Twentieth Annual Conference on Learning Theory (COLT)*, pages 484–498, 2007.

[ACBF02] P. Auer, N. Cesa-Bianchi, and P. Fischer. Finite-time analysis of the multiarmed bandit problem. *Machine Learning Journal*, 47:235–256, 2002.

[ACBFS02] P. Auer, N. Cesa-Bianchi, Y. Freund, and R. Schapire. The nonstochastic multiarmed bandit problem. *SIAM Journal on Computing*, 32:48–77, 2002.

[ACBG02] P. Auer, N. Cesa-Bianchi, and C. Gentile. Adaptive and self-confident on-line learning algorithms. *Journal of Computer and System Sciences*, 64:48–75, 2002.

[Agr95] R. Agrawal. The continuum-armed bandit problem. *SIAM Journal on Control and Optimization*, 33:1926–1951, 1995.

[AHKS06] A. Agarwal, E. Hazan, S. Kale, and R.E. Schapire. Algorithms for portfolio management based on the Newton method. In *Proceedings of the Twenty-Third International Conference on Machine Learning*, 2006.

[ANN04] C. Allenberg-Neeman and B. Neeman. Full information game with gains and losses. In *Proceedings of the Fifteenth International Conference on Algorithmic Learning Theory (ALT)*, pages 264–278, 2004.

[AOS07] P. Auer, R. Ortner, and C. Szepesvári. Improved rates for the stochastic continuum-armed bandit problem. In *Proceedings of the Twentieth Annual Conference on Learning Theory (COLT)*, pages 454–468, 2007.

[APS06] A. Antoniadis, E. Paparoditis, and T. Sapatinas. A functional wavelet–kernel approach for time series prediction. *Journal of the Royal Statistical Society: Series B*, 68(5):837–857, 2006.

[Aum74] R.J. Aumann. Subjectivity and correlation in randomized strategies. *Journal of Mathematical Economics*, 1:67–96, 1974.

[Aum87] R.J. Aumann. Correlated equilibrium as an expression of Bayesian rationality. *Econometrica*, 55:1–18, 1987.

[AW01] K.S. Azoury and M. Warmuth. Relative loss bounds for on-line density estimation with the exponential family of distributions. *Machine Learning*, 43:211–246, 2001.

[BDR05] A. Bruhns, G. Deurveilher, and J.-S. Roy. A non-linear regression model for mid-term load forecasting and improvements in seasonnality. In *Proceedings of the Fifteenth Power Systems Computation Conference (PSCC)*, 2005.

[BEYG00] A. Borodin, R. El-Yaniv, and V. Gogan. On the competitive theory and practice of portfolio selection. In *Proceedings of the Fourth Latin American Symposium on Theoretical Informatics (LATIN)*, pages 173–196, 2000.

[BK96] A.N. Burnetas and M.N. Katehakis. Optimal adaptive policies for sequential allocation problems. *Advances in Applied Mathematics*, 17:122–142, 1996.

[Bla56] D. Blackwell. An analog of the minimax theorem for vector payoffs. *Pacific Journal of Mathematics*, 6:1–8, 1956.

[Blu97] A. Blum. Empirical support for winnow and weighted-majority algorithms: Results on a calendar scheduling domain. *Machine Learning*, 26:5–23, 1997.

[BM07] A. Blum and Y. Mansour. From external to internal regret. *Journal of Machine Learning Research*, 8:1307–1324, 2007.

[Cau01] R. Cauty. Solution du problème de point fixe de Schauder. *Fundamenta Mathematicæ*, 170:231–246, 2001.

[CB99] N. Cesa-Bianchi. Analysis of two gradient-based algorithms for on-line regression. *Journal of Computer and System Sciences*, 59(3):392–411, 1999.

[CBFH+97] N. Cesa-Bianchi, Y. Freund, D. Haussler, D.P. Helmbold, R. Schapire, and M. Warmuth. How to use expert advice. *Journal of the ACM*, 44(3):427–485, 1997.

[CBL03] N. Cesa-Bianchi and G. Lugosi. Potential-based algorithms in on-line prediction and game theory. *Machine Learning*, 51:239–261, 2003.

[CBL06] N. Cesa-Bianchi and G. Lugosi. *Prediction, Learning, and Games*. Cambridge University Press, 2006.

[CM07] P.-A. Coquelin and R. Munos. Bandit algorithms for tree search. In *Proceedings of the Twenty-Third Conference on Uncertainty in Artificial Intelligence (UAI)*, pages 67–74, 2007.

[Cop09] E. Cope. Regret and convergence bounds for immediate-reward reinforcement learning with continuous action spaces. *IEEE Transactions on Automatic Control*, 54(6):1243–1253, 2009.

[Cov65] T. Cover. Behavior of sequential predictors of binary sequences. In *Proceedings of the Fourth Prague Conference on Information Theory, Statistical Decision Functions, Random Processes*, pages 263–272. Czechoslovakian Academy of Sciences, Prague, 1965.

[Cov91] T.M. Cover. Universal portfolios. *Mathematical Finance*, 1:1–29, 1991.

[CW96] X. Chen and H. White. Laws of large numbers for Hilbert space-valued mixingales with applications. *Econometric Theory*, 12(2):284–304, 1996.

[dFM03] D.P. de Farias and N. Megiddo. How to combine expert (or novice) advice when actions impact the environment. In *Proceedings of the Seventeenth Annual Conference on Neural Information Processing Systems (NIPS)*, 2003.

[DGS09] M. Devaine, Y. Goude, and G. Stoltz. Aggregation of sleeping predictors to forecast electricity consumption. Technical report, EDF R&D and École normale supérieure, Paris, August 2009. See `http://www.math.ens.fr/%7stoltz/DeGoSt-report.pdf`.

[DKO+08] V. Dordonnat, S.J. Koopman, M. Ooms, A. Dessertaine, and J. Collet. An hourly periodic state space model for modelling French national electricity load. *International Journal of Forecasting*, 24:566–587, 2008.

[DLS07] O. Dekel, P.M. Long, and Y. Singer. Online learning of multiple tasks with a shared loss. *Journal of Machine Learning Research*, 8:2233–2264, 2007.

[DMP+06] V. Dani, O. Madani, D. Pennock, S. Sanghai, and B. Galebach. An empirical comparison of algorithms for aggregating expert predictions. In *Proceedings of the Twenty-Second Conference on Uncertainty in Artificial Intelligence (UAI)*, 2006.

[EHJT04] B. Efron, T. Hastie, I. Johnstone, and R. Tibshirani. Least angle regression. *Annals of Statistics*, 32(2):407–499, 2004.

[FL99] D. Fudenberg and D. Levine. An easier way to calibrate. *Games and Economic Behavior*, 29:131–137, 1999.

[Fos91] D. Foster. Prediction in the worst-case. *Annals of Statistics*, 19:1084–1090, 1991.

[Fos99] D. Foster. A proof of calibration via Blackwell's approachability theorem. *Games and Economic Behavior*, 29:73–78, 1999.

[FS97] 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, 1997.

[FSSW97] Y. Freund, R. Schapire, Y. Singer, and M. Warmuth. Using and combining predictors that specialize. In *Proceedings of the Twenty-Ninth Annual ACM Symposium on the Theory of Computing (STOC)*, pages 334–343, 1997.

[FV91] D. Foster and R. Vohra. Asymptotic calibration. Technical report, Graduate School of Business, University of Chicago, 1991.

[FV98] D. Foster and R. Vohra. Asumptotic calibration. *Biometrika*, 85:379–390, 1998.

[FV99] D. Foster and R. Vohra. Regret in the on-line decision problem. *Games and Economic Behavior*, 29:7–36, 1999.

[Ger10] S. Gerchinovitz. Personal communication, 2010.

[GLU06] L. Györfi, G. Lugosi, and F. Udina. Nonparametric kernel-based sequential investment strategies. *Mathematical Finance*, 16:337–358, 2006.

[GMS08] S. Gerchinovitz, V. Mallet, and G. Stoltz. A further look at sequential aggregation rules for ozone ensemble forecasting. Technical report, INRIA Paris-Rocquencourt and École normale supérieure, Paris, September 2008. See `http://www.math.ens.fr/%7Estoltz/GeMaSt-report.pdf`.

[GO07] L. Györfi and G. Ottucsák. Sequential prediction of unbounded stationary time series. *IEEE Transactions on Information Theory*, 53(5):1866–1872, 2007.

[Gou08a] Y. Goude. *Mélange de prédicteurs et application à la prévision de consommation électrique*. PhD thesis, Université Paris-Sud, January 2008. The hosting institution was the R&D center of EDF.

[Gou08b] Y. Goude. Tracking the best predictor with a detection based algorithm. In *Proceedings of the Joint Statistical Meetings*. American Statistical Association, 2008. See the "Statistical Computing" section.

[GWMT06] S. Gelly, Y. Wang, R. Munos, and O. Teytaud. Modification of UCT with patterns in Monte-Carlo go. Technical Report RR-6062, INRIA, 2006.

[Han57] J. Hannan. Approximation to Bayes risk in repeated play. In M. Dresher, A. Tucker, and P. Wolfe, editors, *Contributions to the Theory of Games*, volume III, pages 97–139. Princeton University Press, 1957.

[Har95] S. Hart. Personal communication to Dean P. Foster, 1995.

[HK70] A.E. Hoerl and R.W. Kennard. Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, 12:55–67, 1970.

[HK08] E. Hazan and S. Kale. Extracting certainty from uncertainty: Regret bounded by variation in costs. In *Proceedings of the Twenty-First Annual Conference on Learning Theory (COLT)*, 2008.

[HMC00] S. Hart and A. Mas-Colell. A simple adaptive procedure leading to correlated equilibrium. *Econometrica*, 68:1127–1150, 2000.

[HP97] D.P. Helmbold and S. Panizza. Some label efficient learning results. In *Proceedings of the Tenth Annual Conference on Computational Learning Theory (COLT)*, pages 218–230, 1997.

[HS89] S. Hart and D. Schmeidler. Existence of correlated equilibria. *Mathematics of Operations Research*, 14:18–25, 1989.

[HSSW98] D.P. Helmbold, R.E. Schapire, Y. Singer, and M.W. Warmuth. On-line portfolio selection using multiplicative updates. *Mathematical Finance*, 8:325–344, 1998.

[HT10] J. Honda and A. Takemura. An asymptotically optimal bandit algorithm for bounded support models. In *Proceedings of the Twenty-Third Annual Conference on Learning Theory (COLT)*, 2010.

[HW98] M. Herbster and M. Warmuth. Tracking the best expert. *Machine Learning*, 32:151–178, 1998.

[Kle04] R. Kleinberg. Nearly tight bounds for the continuum-armed bandit problem. In *Proceedings of the Eighteenth Annual Conference on Neural Information Processing Systems (NIPS)*, 2004.

[KS06] L. Kocsis and C. Szepesvari. Bandit based Monte-Carlo planning. In *Proceedings of the Fifteenth European Conference on Machine Learning (ECML)*, pages 282–293, 2006.

[KSU08] R. Kleinberg, A. Slivkins, and E. Upfal. Multi-armed bandits in metric spaces. In *Proceedings of the Fortieth Annual ACM Symposium on the Theory of Computing (STOC)*, 2008.

[KV03] A. Kalai and S. Vempala. Efficient algorithms for the online decision problem. In *Proceedings of the Sixteenth Annual Conference on Learning Theory (COLT)*, pages 26–40. Springer, 2003.

[KW97] J. Kivinen and M. Warmuth. Exponentiated gradient versus gradient descent for linear predictors. *Information and Computation*, 132(1):1–63, 1997.

[LR85] T.L. Lai and H. Robbins. Asymptotically efficient adaptive allocation rules. *Advances in Applied Mathematics*, 6:4–22, 1985.

[LS07] E. Lehrer and E. Solan. Learning to play partially-specified equilibrium. Submitted for publication, 2007.

[LW94] N. Littlestone and M. Warmuth. The weighted majority algorithm. *Information and Computation*, 108:212–261, 1994.

[LZ76] A. Lempel and J. Ziv. On the complexity of an individual sequence. *IEEE Transactions on Information Theory*, 22:75–81, 1976.

[Mal10] V. Mallet. Ensemble forecast of analyses: Coupling data assimilation and sequential aggregation. Submitted for publication, 2010.

[MMS07] V. Mallet, B. Mauricette, and G. Stoltz. Description of sequential aggregation methods and their performance for ozone ensemble forecasting. Technical Report DMA-07-08, École normale supérieure, Paris, 2007.

[MS03] S. Mannor and N. Shimkin. On-line learning with imperfect monitoring. In *Proceedings of the Sixteenth Annual Conference on Learning Theory*, pages 552–567. Springer, 2003.

[MS06] V. Mallet and B. Sportisse. Ensemble-based air quality forecasts: A multimodel approach applied to ozone. *Journal of Geophysical Research*, 111(D18), 2006.

[MSZ94] J.-F. Mertens, S. Sorin, and S. Zamir. Repeated games. Technical Report 9420, 9421, 9422, Université catholique de Louvain, 1994.

[Per09a] V. Perchet. Approachability of convex sets in games with partial monitoring. Submitted for publication, 2009.

[Per09b] V. Perchet. Calibration and internal no-regret with random signals. In *Proceedings of the Twentieth International Conference on Algorithmic Learning Theory (ALT)*, pages 68–82, 2009.

[Per09c] V. Perchet. No-regret with partial monitoring calibration-based optimal algorithms. Submitted for publication, 2009.

[Per10] V. Perchet. *Approchabilité, calibration et regret dans les jeux à observations partielles*. PhD thesis, Université Paris VI Pierre-et-Marie-Curie, 2010.

[PLG09] A. Pierrot, N. Laluque, and Y. Goude. Short-term electricity load forecasting with generalized additive models. In *Proceedings of the Third International Conference on Computational and Financial Econometrics (CFE)*, 2009.

[PS01] A. Piccolboni and C. Schindelhauer. Discrete prediction games with arbitrary feedback and loss. In *Proceedings of the Fourteenth Annual Conference on Computational Learning Theory*, pages 208–223, 2001.

[Rob52] H. Robbins. Some aspects of the sequential design of experiments. *Bulletin of the American Mathematics Society*, 58:527–535, 1952.

[Rus99] A. Rustichini. Minimizing regret: The general case. *Games and Economic Behavior*, 29:224–243, 1999.

[Sto05] G. Stoltz. *Information incomplète et regret interne en prédiction de suites individuelles*. PhD thesis, Université Paris-Sud, May 2005.

[Tib96] R. Tibshirani. Regression shrinkage and selection via the Lasso. *Journal of the Royal Statistical Society, Series B*, 58(1):267–288, 1996.

[Vov90] V. Vovk. Aggregating strategies. In *Proceedings of the Third Annual Workshop on Computational Learning Theory (COLT)*, pages 372–383, 1990.

[Vov98] V. Vovk. A game of prediction with expert advice. *Journal of Computer and System Sciences*, 56(2):153–173, 1998.

[Vov01] V. Vovk. Competitive on-line statistics. *International Statistical Review*, 69:213–248, 2001.

[VZ08] V. Vovk and F. Zhdanov. Prediction with expert advice for the Brier game. In *Proceedings of the Twenty-Fifth International Conference on Machine Learning (ICML)*, 2008.

[Woo06] S.N. Wood. *Generalized Additive Models: An Introduction with R*. Chapman and Hall/CRC, 2006.

[Ziv78] J. Ziv. Coding theorems for individual sequences. *IEEE Transactions on Information Theory*, 24:405–412, 1978.

[Ziv80] J. Ziv. Distortion-rate theory for individual sequences. *IEEE Transactions on Information Theory*, 26:137–143, 1980.

[ZL77] J. Ziv and A. Lempel. A universal algorithm for sequential data-compression. *IEEE Transactions on Information Theory*, 23:337–343, 1977.