On the optimality of the Hedge algorithm in the stochastic regime

Jaouad Mourtada* Stéphane Gaïffas†

December 22, 2018

Abstract

In this paper, we study the behavior of the Hedge algorithm in the online stochastic setting. We prove that anytime Hedge with decreasing learning rate, which is one of the simplest algorithm for the problem of prediction with expert advice, is surprisingly both worst-case optimal and adaptive to the easier stochastic and adversarial with a gap problems. This shows that, in spite of its small, non-adaptive learning rate, Hedge possesses the same optimal regret guarantee in the stochastic case as recently introduced adaptive algorithms [7, 9, 11, 14, 17]. Moreover, our analysis exhibits qualitative differences with other variants of the Hedge algorithm, such as the fixed-horizon version (with constant learning rate) and the one based on the so-called "doubling trick", both of which fail to adapt to the easier stochastic setting. Finally, we discuss the limitations of anytime Hedge and the improvements provided by second-order regret bounds in the stochastic case.

Keywords. Online learning; prediction with expert advice; Hedge; adaptive regret.

1 Introduction

The standard setting of *prediction with expert advice* [16, 10, 27, 6] aims to provide sound strategies for sequential prediction that combine the forecasts from different sources. More precisely, in the so-called *Hedge problem* [10], at each round the learner has to output a probability distribution on a finite set of *experts* $\{1, \ldots, M\}$; the losses of the experts are then revealed, and the learner incurs the expected loss from its chosen probability distribution. The goal is then to control the *regret*, defined as the difference between the cumulative loss of the learner and that of the best expert (with smallest loss). This online prediction problem is typically considered in the *individual sequences* framework, where the losses may be arbitrary and in fact set by an adversary that seeks to maximize the regret. This leads to regret bounds that hold under virtually no assumption [6].

In this setting, arguably the simplest and most standard strategy is the *Hedge algorithm* [10], also called the *exponentially weighted averaged forecaster* [6]. This algorithm depends on a time-varying parameter η_t called the *learning rate*, which quantifies by how much the algorithm departs from its initial probability distribution to put more weight on the currently leading experts. Given a known finite time horizon T, the standard tuning of the learning rate is fixed and given by $\eta_t = \eta \propto \sqrt{\log(M)/T}$, which guarantees an optimal worst-case regret of order $O(\sqrt{T \log M})$. Alternatively, when T is unknown, one can set $\eta_t \propto \sqrt{\log(M)/t}$ at round t, which leads to an *anytime* $O(\sqrt{T \log M})$ regret bound valid for all $T \geqslant 1$.

[*] Centre de Mathématiques Appliquées, École Polytechnique, Palaiseau, France

[†] LPSM, UMR 8001, Université Paris Diderot, Paris, France

Related work. While worst-case regret bounds are robust and always valid, they turn out to be overly pessimistic. A recent line of research [7, 9, 11, 15, 21, 14, 17] designs algorithms that combine the worst-case $O(\sqrt{T \log M})$ regret guarantees with an improved regret on easier instances of the problem. An interesting example of such an easier instance is the stochastic problem, where it is assumed that the losses are stochastic and that at each round the expected loss of a "leader" expert is smaller than those of the other experts by some gap Δ. Such algorithms rely either on a more careful, data-dependent tuning of the learning rate η_t [7, 9, 15, 11], or on more sophisticated strategies [14, 17]. As shown in [11] (see also [13]), one particular type of adaptive regret bounds (so-called *second-order bounds*) implies at the same time a $O(\sqrt{T \log M})$ worst-case bound and a better *constant* $O(\log(M)/\Delta)$ bound in the stochastic problem with gap Δ. Arguably starting with the early work on second-order bounds [7], the design of online learning algorithms that combine robust worst-case guarantees with improved performance on easier instances has been an active research goal in recent years [9, 11, 15, 21]. However, to the best of our knowledge, the existing work has focused on developing new adaptive algorithms rather than on analyzing the behavior of "conservative" algorithms in favorable scenarios. Owing to the fact that the standard Hedge algorithm is designed for — and analyzed in — the adversarial setting [16, 10, 6], and that its parameters are not tuned adaptively to obtain better bounds in easier instances, it may be considered as overly conservative and not adapted to stochastic environments.

Our contribution. This paper fills a gap in the existing literature by providing an analysis of the standard Hedge algorithm in the stochastic setting. We show that, quite surprisingly, the anytime Hedge algorithm with default learning rate $\eta_t \propto \sqrt{\log(M)/t}$ actually *adapts* to the stochastic setting, in which it achieves an optimal *constant* $O(\log(M)/\Delta)$ regret bound *without any particular dedicated tuning* for the easier instance. This contrasts with previous works, which require the construction of new adaptive (and more involved) algorithms. Remarkably, this property is *not* shared by the variant of Hedge for a known fixed-horizon T with constant learning rate $\eta \propto \sqrt{\log(M)/T}$, since it suffers a $\Theta(\sqrt{T \log M})$ regret even in easier instances. This explicits a very strong difference between the performances of the anytime and the fixed-horizon variants of the Hedge algorithm.

Outline. We define the setting of prediction with expert advice and the Hedge algorithm in Section 2, and we recall herein its standard worst-case regret bound. In Section 3, we consider the behavior of the Hedge algorithm on easier instances, namely the stochastic setting with a gap Δ on the best expert. Under an i.i.d assumption on the sequence of losses, we provide in Theorem 1 an upper bound on the regret of order $(\log M)/\Delta$ for Decreasing Hedge. In Proposition 2, we prove that the rate $(\log M)/\Delta$ cannot be improved in this setting. In Theorem 2 and Corollary 1, we extend the regret guarantees to the adversarial with a gap setting, where a leading expert linearly outperforms the others. These results stand for any Hedge algorithm which is worst-case optimal and with any learning rate which is larger than the one of Decreasing Hedge, namely $O(\sqrt{\log M)/t})$. In Proposition 3, we prove the sub-optimality of the fixed-horizon Hedge algorithm, and of another version of Hedge based on the so-called "doubling trick". In Section 4, we discuss the advantages of adaptive versions of Hedge, and explain what are the limits of Decreasing Hedge compared to such versions. We provide numerical illustrations of our theoretical findings in Section 5, conclude in Section 6 and proofs are given in Section 7.

2 The expert problem and the Hedge algorithm

In the Hedge setting, also called *decision-theoretic online learning* (DTOL) [10], the learner and its adversary (the Environment) sequentially compete on the following game: at each round $t \geqslant 1$,

1. the Learner chooses a probability vector $\boldsymbol{v}_t = (v_{i,t})_{1 \leqslant i \leqslant M}$ on the M experts $1, \ldots, M$;
2. the Environment picks a bounded loss vector $\boldsymbol{\ell}_t = (\ell_{i,t})_{1 \leqslant i \leqslant M} \in [0,1]^M$, where $\ell_{i,t}$ is the loss of expert i at round t, while the Learner suffers loss $\ell_t = \boldsymbol{v}_t^\top \boldsymbol{\ell}_t$.

The goal of the Learner is to control its *regret*

$$R_T = \sum_{t=1}^T \ell_t - \min_{1 \leqslant i \leqslant M} \sum_{t=1}^T \ell_{i,t} \tag{1}$$

for every $T \geqslant 1$, irrespective of the sequence of loss vectors $\boldsymbol{\ell}_1, \boldsymbol{\ell}_2, \ldots$ chosen by the Environment. One of the most standard algorithms for this setting is the *Hedge* algorithm. The Hedge algorithm, also called the exponentially weighted averaged forecaster, uses the vector of probabilities $\boldsymbol{v}_t = (v_{i,t})_{1 \leqslant i \leqslant M}$ given by

$$v_{i,t} = \frac{e^{-\eta_t L_{i,t-1}}}{\sum_{j=1}^M e^{-\eta_t L_{j,t-1}}} \tag{2}$$

at each $t \geqslant 1$, where $L_{i,T} = \sum_{t=1}^T \ell_{i,t}$ denotes the cumulative loss of expert i for every $T \geqslant 1$. Let us also denote $L_T := \sum_{t=1}^T \ell_t$. We consider in this paper the following variants of Hedge.

Decreasing Hedge [1]. This is Hedge with the sequence of learning rates $\eta_t \propto \sqrt{\log(M)/t}$.

Constant Hedge [16]. Given a finite time horizon $T \geqslant 1$, this is Hedge with constant learning rate $\eta_t = \eta \propto \sqrt{\log(M)/T}$.

Hedge with doubling trick [5, 6]. This variant of Hedge uses a constant learning rate on geometrically increasing intervals, restarting the algorithm at the beginning of each interval. Namely, it uses

$$v_{i,t} = \frac{\exp(-\eta_t \sum_{s=T_k}^{t-1} \ell_{i,s})}{\sum_{j=1}^M \exp(-\eta_t \sum_{s=T_k}^{t-1} \ell_{j,s})}, \tag{3}$$

with $T_l = 2^l$ for $l \geqslant 0$, $k \in \mathbf{N}$ such that $T_k \leqslant t < T_{k+1}$ and $\eta_t = c_0 \sqrt{\log(M)/T_k}$, with $c_0 > 0$.

Let us recall the following standard regret bound for the Hedge algorithm from [8].

Proposition 1. *Let $\eta_1, \eta_2, \ldots$ be a decreasing sequence of learning rates. The Hedge algorithm* (2) *satisfies the following regret bound:*

$$R_T \leqslant \frac{1}{\eta_T} \log M + \frac{1}{8} \sum_{t=1}^T \eta_t \,. \tag{4}$$

In particular, the choice $\eta_t = 2\sqrt{\log(M)/t}$ yields a regret bound of $\sqrt{T \log M}$ for every $T \geqslant 1$.

Note that the regret bound stated in Equation (4) holds for every sequence of losses $\boldsymbol{\ell}_1, \boldsymbol{\ell}_2, \dots$, which makes it valid under no assumption (aside from the boundedness of the losses). The worst-case regret bound in $O(\sqrt{T \log M})$ is achieved by Decreasing Hedge, Hedge with doubling trick and Constant Hedge (whenever T is known in advance). The $O(\sqrt{T \log M})$ rate cannot be improved either by Hedge or any other algorithm: it is known to be the minimax optimal regret [6]. Contrary to Constant Hedge, Decreasing Hedge is anytime, in the sense that it achieves the $O(\sqrt{T \log M})$ regret bound simultaneously for each $T \geqslant 1$. We note that this worst-case regret analysis fails to exhibit any difference between these three algorithms.

In many cases, this $\sqrt{T}$ regret bound is pessimistic, and more "aggressive" strategies (such as the follow-the-leader algorithm, which plays at each round the uniform distribution on the experts with smallest loss [6]) may achieve constant regret in easier instances, even though they lack regret guarantees in the adversarial regime. We show in Section 3 below that Decreasing Hedge is actually better than both Constant Hedge and Hedge with doubling trick in some easier instance of the problem (including in the stochastic setting). This entails that Decreasing Hedge is actually able to adapt, without any modification, to the easiness of the problem considered.

3 Regret bounds for Hedge variants on easy instances

In this Section, we depart from the worst-case regret analysis and study the regret of the considered variants of the Hedge algorithm on easier instances of the prediction with expert advice problem.

3.1 Optimal regret for Decreasing Hedge in the stochastic regime

We examine the behavior of Decreasing Hedge in the stochastic regime, where the losses are the realization of some (unknown) stochastic process. More precisely, we consider the standard i.i.d. case, where the loss vectors $\boldsymbol{\ell}_1, \boldsymbol{\ell}_2, \dots$ are i.i.d. (independence holds over rounds, but not necessarily across experts). In this setting, the regret can be much smaller than the worst-case $\sqrt{T \log M}$ regret, since the best expert (with smallest expected loss) will dominate the rest after some time. Following [11, 17], the easiness parameter we consider in this case, which governs the time needed for the best expert to have the smallest cumulative loss and hence the incurred regret, is the sub-optimality gap $\Delta = \min_{i \neq i^*} \mathbb{E}[\ell_{i,t} - \ell_{i^*,t}]$.

We show below that, despite the fact that Decreasing Hedge is designed for the worst-case setting described in Section 2, it is able to adapt to the easier problem considered here, Indeed, Theorem 1 shows that Decreasing Hedge achieves a *constant*, and in fact *optimal* (by Proposition 2 below) regret bound in this setting, in spite of its "conservative" learning rate.

Theorem 1. *Let $M \geqslant 3$. Assume that the loss vectors $\boldsymbol{\ell}_1, \boldsymbol{\ell}_2, \dots$ are i.i.d. random variables, where $\boldsymbol{\ell}_t = (\ell_{i,t})_{1 \leqslant i \leqslant M}$. Also, assume that there exists $i^* \in \{1, \dots, M\}$ and $\Delta > 0$ such that*

$$\mathbb{E}[\ell_{i,t} - \ell_{i^*,t}] \geqslant \Delta \tag{5}$$

for every $i \neq i^*$. Then, the Decreasing Hedge algorithm with learning rate $\eta_t = 2\sqrt{(\log M)/t}$ achieves the following regret bound: for every $T \geqslant 1$,

$$\mathbb{E}[R_T] \leqslant \frac{4 \log M + 27}{\Delta} \,. \tag{6}$$

The proof of Theorem 1 is given in Section 7.1.1. Theorem 1 proves that, in the stochastic setting with a gap Δ, the Decreasing Hedge algorithm achieves a regret $O(\log(M)/\Delta)$, without any prior knowledge of Δ. This matches the guarantees of adaptive Hedge algorithms which are explicitly designed to adapt to easier instances [11, 17]. This result may seem surprising at first: indeed, adaptive exponential weights algorithms that combine optimal regret in the adversarial setting and constant regret in easier scenarios, such as Hedge with a second-order tuning [7] or AdaHedge [9], typically use a data-dependent learning rate η_t that adapts to the properties of the losses. While the learning rate η_t chosen by these algorithms may be as low as the worst-case tuning $\eta_t \propto \sqrt{\log(M)/t}$, in the stochastic case those algorithms will use larger, lower-bounded learning rates to ensure constant regret. As Theorem 1 above shows, it turns out that the data-independent, "safe" learning rates $\eta_t \propto \sqrt{\log(M)/t}$ used by "vanilla" Decreasing Hedge are still large enough to adapt to the stochastic case.

Idea of the proof. The idea of the proof of Theorem 1 is to divide time in two phases: a short initial phase $[\![1, t_1]\!]$, where $t_1 = O(\frac{\log M}{\Delta^2})$, and a second phase $[\![t_1, T]\!]$. The initial phase is dominated by noise, and regret during this period is bounded through the worst-case regret bound of Proposition 1, which gives a regret of $O(\sqrt{t_1 \log M}) = O(\frac{\log M}{\Delta})$. In the second phase, the best expert dominates the rest, and the weights concentrate on this best expert fast enough that the total regret incurred is small. The control of the regret in the second phase relies on the critical fact that, if η_t is at least as large as $\sqrt{(\log M)/t}$, then the following two things occur simultaneously at $t_1 \asymp \frac{\log M}{\Delta^2}$, namely at the beginning of the late phase:

1. the best expert i^* dominates all the others linearly: for every $i \neq i^*$ and $t \geqslant t_1$, $L_{i,t} - L_{i^*,t} \geqslant \frac{\Delta t}{2}$;
2. the total weight of all suboptimal experts is controlled: $\sum_{i \neq i^*} v_{i,t_1} \leqslant \frac{1}{2}$. If $\eta_t \geqslant \sqrt{(\log M)/t}$ and the first condition holds, this amounts to $M \exp(-\frac{\Delta}{2}\sqrt{t \log M}) \leqslant \frac{1}{2}$, namely $t_1 \gtrsim \frac{\log M}{\Delta^2}$.

In other words, $\eta_t \asymp \sqrt{(\log M)/t}$ is the minimal learning rate which ensures that the total weight of suboptimal experts starts vanishing at about the same time as when the best expert starts to dominate the others with a large probability (and remarkably, this property holds for every value of the sub-optimality gap Δ). Finally, the upper bound on the regret in the second phase rests on the two conditions above, together with the bound $\sum_{t \geqslant 0} e^{-c\sqrt{t}} = O(\frac{1}{c^2})$ for $c > 0$.

Remark 1. The fact that $\sum_{t \geqslant 0} e^{-c\sqrt{t}} = O(\frac{1}{c^2})$ is used in the analysis of the EXP3++ bandit algorithm [23] (Lemma 10 herein), which combines a $O(\sqrt{MT \log M})$ regret bound in the adversarial setting with the $O(M \log^2(T)/\Delta)$ regret bound in the i.i.d. stochastic setting with gap Δ (where M denotes the number of arms) [22]. Note however that summing the contribution of all experts (or arms), which suffices in the bandit setting to obtain the optimal order of regret, would yield a significantly suboptimal $O(\frac{M}{\Delta})$ regret bound in the expert setting considered here, where optimal dependence w.r.t. the number of experts is $\log M$. In our case, the decomposition of the regret in two phases, which is explained above, removes the linear dependence on M and allows to achieve the optimal rate $(\log M)/\Delta$.

We complement Theorem 1 by showing that the $O((\log M)/\Delta)$ regret under the gap condition cannot be improved, in the sense that its dependence on both M and Δ is optimal. To the best of our knowledge, such a lower-bound was not previously given in the literature in this setting.

Proposition 2. *Let $\Delta \in (0, \frac{1}{6})$, $M \geqslant 2$ and $T \geqslant (\log M)/(18\Delta^2)$. Then, for any Hedge algorithm, there exists an i.i.d. distribution over the sequence of losses $(\boldsymbol{\ell}_t)_{t \geqslant 1}$ such that:*

- *there exists $i^* \in \{1, \dots, M\}$ such that, for any $i \neq i^*$, $\mathbb{E}[\ell_{i,t} - \ell_{i^*,t}] \geqslant \Delta$;*
- *the expected regret of the algorithm satisfies:*

$$\mathbb{E}[R_T] \geqslant \frac{\log M}{620\Delta}\,. \tag{7}$$

The proof of Proposition 2 is given in Section 7.2.3.

3.2 Small regret for Decreasing Hedge in the adversarial with a gap problem

In this section, we extend the regret guarantee of Decreasing Hedge in the stochastic setting (Theorem 1), by showing that it holds for more general algorithms and under more general assumptions. Specifically, we consider an "adversarial with a gap" regime, similar to the one introduced in [23] in the bandit case, where the leading expert linearly outperforms the others after some time. As Theorem 2 shows, essentially the same regret guarantee can be obtained in this case, up to an additional $\log(\Delta^{-1})/\Delta$ term. Theorem 2 also applies to any Hedge algorithm whose (possibly data-dependent) learning rate η_t is at least as large as that of Decreasing Hedge, and which satisfies a $O(\sqrt{T \log M})$ worst-case regret bound; this includes algorithms with *anytime* first and second-order tuning of the learning rate [1, 7, 9].

Theorem 2. *Let $M \geqslant 3$. Assume that there exists $\tau_0 \geqslant 1$, $\Delta > 0$ and $i^* \in \{1, \dots, M\}$ such that, for every $t \geqslant \tau_0$ and $i \neq i^*$, one has*

$$L_{i,t} - L_{i^*,t} \geqslant \Delta t. \tag{8}$$

Consider any Hedge algorithm with (possibly data-dependent) learning rate η_t such that

- *$\eta_t \geqslant c_0 \sqrt{(\log M)/t}$ for some constant $c_0 > 0$;*
- *the Hedge algorithm with learning rate η_t admits the following worst-case regret bound: $R_T \leqslant c_1 \sqrt{T \log M}$ for every $T \geqslant 1$, for some $c_1 > 0$.*

Then, for every $T \geqslant 1$, the regret of this algorithm is upper bounded as

$$R_T \leqslant c_1 \sqrt{\tau_0 \log M} + \frac{c_2 \log M + c_3 \log \Delta^{-1} + c_4}{\Delta} \tag{9}$$

where $c_2 = c_1 + \frac{\sqrt{8}}{c_0}$, $c_3 = \frac{\sqrt{8}}{c_0}$ and $c_4 = \frac{16}{c_0^2}$.

The idea of the proof of Theorem 2 is the same as that of Theorem 1, the only difference being the slightly longer initial phase to account for the adversarial nature of the losses. As a consequence of the general bound of Theorem 2, we can recover the guarantee of Theorem 1 (up to an additional $\log(\Delta^{-1})/\Delta$ term), both in expectation and with high probability, under more general stochastic assumptions than i.i.d. over time. The proofs of Theorem 2 and Corollary 1 are provided in Section 7.1.2.

Corollary 1. *Assume that the losses* $(\ell_{i,t})_{1 \leqslant i \leqslant M, t \geqslant 1}$ *are random variables. Also, denoting* $\mathcal{F}_t = \sigma\big((\ell_{i,s})_{1 \leqslant i \leqslant M, 1 \leqslant s \leqslant t}\big)$, *assume that there exists* i^* *and* $\Delta > 0$ *such that*

$$\mathbb{E}\left[\ell_{i,t} - \ell_{i^*,t} \,|\, \mathcal{F}_{t-1}\right] \geqslant \Delta \tag{10}$$

for every $i \neq i^*$ *and every* $t \geqslant 1$. *Then, for any Hedge algorithm satisfying the conditions of Theorem 2, and every* $T \geqslant 1$:

$$\mathbb{E}[R_T] \leqslant (5c_1 + 2c_2)\frac{\log M}{\Delta} + 2c_3\frac{\log \Delta^{-1}}{\Delta} + \frac{2c_4}{\Delta}, \tag{11}$$

with c_1, c_2, c_3, c_4 *as in Theorem 2. In addition, for every* $\varepsilon \in (0,1)$, *we have*

$$R_T \leqslant \left(c_1\sqrt{8} + 2c_2\right)\frac{\log M}{\Delta} + c_1\frac{\sqrt{8\log M \log \varepsilon^{-1}}}{\Delta} + 2c_3\frac{\log \Delta^{-1}}{\Delta} + \frac{2c_4}{\Delta} \tag{12}$$

with probability at least $1 - \varepsilon$.

3.3 Constant Hedge and Hedge with the doubling trick do not adapt to the stochastic case

Now, we show that the adaptivity of Decreasing Hedge to gaps in the losses, established in Sections 3.1 and 3.2, is not shared by the two closely related Constant Hedge and Hedge with the doubling trick, despite the fact that they both achieve the minimax optimal worst-case $O(\sqrt{T \log M})$ regret. Proposition 3 below shows that both algorithms fail to achieve a constant regret, and in fact to improve over their worst-case $\Theta(\sqrt{T \log M})$ regret guarantee, even in the extreme case of experts with constant losses 0 (for the leader), and 1 for the rest (*i.e.*, $\Delta = 1$).

Proposition 3. *Let* $T \geqslant 1$, $M \geqslant 2$, *and consider the experts* $i = 1, \dots, M$ *with losses* $\ell_{1,t} = 0$, $\ell_{i,t} = 1$ $(1 \leqslant t \leqslant T, 2 \leqslant i \leqslant M)$. *Then, the regret of Constant Hedge with learning rate* $\eta_t = c_0\sqrt{\log(M)/T}$ (*where* $c_0 > 0$ *is a numerical constant*) *is lower bounded as follows*:

$$R_T \geqslant \min\left(\frac{\sqrt{T \log M}}{3c_0}, \frac{T}{3}\right). \tag{13}$$

In addition, Hedge with doubling trick (3) *also suffers a regret satisfying*

$$R_T \geqslant \min\left(\frac{\sqrt{T \log M}}{6c_0}, \frac{T}{12}\right). \tag{14}$$

The proof of Proposition 3 is given in Section 7.2.1. Although Hedge with a doubling trick is typically considered as overly conservative and only suitable for worst-case scenarios [6] (especially due to its periodic restarts, after which it discards past observations), to the best of our knowledge Proposition 3 (together with Theorem 1) is the first to formally demonstrate the advantage of Decreasing Hedge over the doubling trick version. This implies that Decreasing Hedge should not be seen as merely a substitute for Constant Hedge to achieve anytime regret bounds. Indeed, even when the horizon T is fixed, Decreasing Hedge outperforms Constant Hedge in the stochastic setting.

4 The advantage of second-order algorithms in the stochastic case

Besides the vanilla variants of Hedge discussed in the previous sections (Constant Hedge, Decreasing Hedge and Hedge with doubling trick), several adaptive Hedge algorithms have been introduced in the literature in recent years.

A first improvement is given by *first-order bounds* [5, 1, 6], where the regret scales as $O(\sqrt{L_T^* \log M} + \log M)$, where L_T^* denotes the cumulative loss of the best expert. This improvement is valid in the general adversarial case, and can be achieved without the knowledge of L_T^* (either through the doubling trick, or by tuning the learning rate as $\eta_t \approx \sqrt{(\log M)/L_{t-1}^*}$. The Decreasing Hedge algorithm, which is invariant under translation of the losses, does not achieve such first order bounds. On the other hand, in a typical stochastic instance (where the best expert has a positive expected loss), the cumulative loss of the best expert grows linearly with T, so that the first-order regret bound still scales as $O(\sqrt{T \log M})$, albeit with a possibly improved leading constant.

A refinement over the first-order bounds is provided by so-called second-order bounds [7, 9, 11, 14, 28], which are algorithm-dependent regret bounds that depend on some notion of variance of the losses across experts (see Equation (16) below for such a bound). While second-order bounds can be hard to interpret, since they typically depend on the weights chosen by the algorithm, they are a convenient notion of adaptivity to easy instances. Indeed, second-order bounds imply both a first-order bound in the adversarial setting, and a constant $O((\log M)/\Delta)$ regret bound in the stochastic case with gap Δ, see [11]. While Decreasing Hedge already achieves the $O((\log M)/\Delta)$ regret bound, adaptive Hedge algorithms with second-order regret bounds do improve over Hedge in some easy stochastic instances, under some additional condition on the losses.

Definition 1 (Bernstein condition)**.** Assume that the losses $\boldsymbol{\ell}_1, \boldsymbol{\ell}_2, \dots$ are the realization of a stochastic process. Denote $\mathcal{F}_t = \sigma(\boldsymbol{\ell}_1, \dots, \boldsymbol{\ell}_t)$ the σ-algebra generated by $\boldsymbol{\ell}_1, \dots, \boldsymbol{\ell}_t$. For $\beta \in [0, 1]$ and $B > 0$, the losses are said to satisfy the (β, B)*-Bernstein condition* if there exists i^* such that, for every $t \geqslant 1$ and $i \neq i^*$,

$$\mathbb{E}[(\ell_{i,t} - \ell_{i^*,t})^2 \,|\, \mathcal{F}_{t-1}] \leqslant B \mathbb{E}[\ell_{i,t} - \ell_{i^*,t} \,|\, \mathcal{F}_{t-1}]^\beta \,. \tag{15}$$

The Bernstein condition [3], a generalization of the Tsybakov margin condition [24, 18], is a geometric property on the losses which enables to obtain fast rates (e.g., faster than $O(1/\sqrt{n})$ for parametric classes) in statistical learning; we refer to [26] for a discussion of fast rates conditions. The Bernstein condition (15) is a measure of the "easiness" of a stochastic instance, which can be seen as a way to generalize the gap condition considered in the previous Section (see Example 1 below). Roughly speaking, it states that good experts (with near-optimal expected loss) are highly correlated with the best expert. As shown by [13] (and implicitly used by [11]), algorithms with second-order regret bounds enjoy improved regret bounds under the Bernstein condition. For completeness, we state this fact in Proposition 4 below. The difference with Theorem 3 in [13] is that we make the dependence on B explicit; for simplicity, we only provide a bound in expectation. The proof of Proposition 4, which uses the same ideas as Theorem 11 in [11], is provided in Section 7.1.3.

Proposition 4. *Consider an algorithm for the Hedge problem which satisfies the following regret bound: for every $i \in \{1, \dots, M\}$, denoting $R_{i,T} := L_T - L_{i,T}$ the regret with respect to expert i,*

$$R_{i,T} \leqslant C_1 \sqrt{(\log M) \sum_{t=1}^{T} (\ell_t - \ell_{i,t})^2 + C_2 \log M} \tag{16}$$

where ℓ_t is the loss incurred by the algorithm, and $C_1, C_2 > 0$ are constants. Assume that the losses satisfy the (β, B)-Bernstein condition. Then, the expected (pseudo-)regret of the algorithm satisfies:

$$\mathbb{E}[R_{i^*,T}] \leqslant C_3 (B \log M)^{\frac{1}{2-\beta}} T^{\frac{1-\beta}{2-\beta}} + C_4 \log M \tag{17}$$

where $C_3 = \max(1, 4C_1^2)$ and $C_4 = 2C_2$.

The data-dependent regret bound (16), a "second-order" bound, is satisfied by adaptive algorithms such as Adapt-ML-Prod [11] and Squint [14].

Example 1. For i.i.d. loss vectors $\boldsymbol{\ell}_1, \boldsymbol{\ell}_2, \dots$, we provide a list of examples.

1. If $\mathbb{E}[\ell_{i,t} - \ell_{i^*,t}] \geqslant \Delta$ for $i \neq i^*$, then the $(1, \frac{1}{\Delta})$-Bernstein condition holds (Lemma 4 in [13]). By Proposition 4, algorithms which satisfy the bound (16) achieve $O(\frac{\log M}{\Delta})$ regret in this case.

2. If $|\ell_{i,t} - \ell_{j,t}| \leqslant \varepsilon$ almost surely for every i, j, then the $(0, \varepsilon^2)$-Bernstein condition holds, and the regret bound (17) becomes $O(\varepsilon\sqrt{T \log M} + \log M)$, which is akin to the Hedge regret bound but with a dependence only on the true range ε of losses.

3. Assuming that $\Delta_i := \mathbb{E}[\ell_{i,t} - \ell_{i^*,t}] \geqslant \Delta$ for $i \neq i^*$, and that the best expert has expected loss $\alpha = \mathbb{E}[\ell_{i^*,t}]$, the $(1, 1 + \frac{2\alpha}{\Delta})$-Bernstein condition is satisfied. Indeed, for any $i \neq i^*$, denoting $\mu_i := \mathbb{E}[\ell_{i,t}] = \alpha + \Delta_i$, we have (since $(u-v)^2 \leqslant \max(u^2, v^2) \leqslant u^2 + v^2 \leqslant u + v$ for $u, v \in [0, 1]$):

$$\mathbb{E}\big[(\ell_{i,t} - \ell_{i^*,t})^2\big] \leqslant \mathbb{E}\left[\ell_{i,t} + \ell_{i^*,t}\right] = \frac{\mu_i + \alpha}{\mu_i - \alpha} \mathbb{E}\left[\ell_{i,t} - \ell_{i^*,t}\right] = \left(1 + \frac{2\alpha}{\Delta_i}\right) \mathbb{E}\left[\ell_{i,t} - \ell_{i^*,t}\right] ,$$

   which establishes the claim since $\Delta_i \geqslant \Delta$. This example combines the small loss and the gap improvement; in this case, by Proposition 4, any algorithm with a second-order bound achieves $O\left((1 + \frac{\alpha}{\Delta}) \log M\right)$ regret. This improves over both the $O(\sqrt{\alpha T \log M})$ first-order bound and Decreasing Hedge's $O(\frac{\log M}{\Delta})$ regret bound.

4. Let P be a distribution on $\mathcal{X} \times \{0, 1\}$, where $\mathcal{X}$ is some measurable space. Assume that $(X_1, Y_1), (X_2, Y_2) \dots$ are i.i.d. samples from P, and that the experts $i \in \{1, \dots, M\}$ correspond to classifiers $f_i : \mathcal{X} \to \{0, 1\}$: $\ell_{i,t} = \mathbf{1}(f_i(X_t) \neq Y_t)$, and that the expert i^* corresponds to the Bayes classifier: $f_{i^*}(X) = \mathbf{1}(\eta(X) \geqslant 1/2)$ almost surely, where $\eta(X) = \mathbb{P}(Y = 1 \,|\, X)$.

   - Tsybakov's low noise condition [24], namely $\mathbb{P}(|2\eta(X) - 1| \leqslant t) \leqslant Ct^\kappa$ for some $C > 0$, $\kappa \geqslant 0$ and every $t > 0$, implies the $(\frac{\kappa}{\kappa+1}, B)$-Bernstein condition for some B (see, e.g., [4]).
   - Under the Massart condition [19] that η is bounded away from $1/2$, namely $|\eta(X) - 1/2| \geqslant c$ almost surely, the $(1, 1/c)$-Bernstein condition holds, so that the regret bound (17) becomes $O(c^{-1} \log M)$.

   Note that these conditions may hold even with an arbitrarily small sub-optimality gap Δ.

Note that in Proposition 4, the (β, B)-Bernstein condition ensures constant expected regret if and only if $\beta = 1$, in which case the regret bound (17) is $O(B \log M)$; in particular, if $\beta < 1$, the bound (17) is asymptotically larger than $O(\frac{\log M}{\Delta})$, where Δ is the sub-optimality gap[1]. We

[1] Note that the Bernstein condition with $\beta > 0$ implies the uniqueness up to duplication of the best expert (with smallest expected loss), and hence the positiveness of Δ.

therefore focus on the case $\beta = 1$. As noted in Example 1, the existence of a gap $\Delta > 0$ implies the $(1, B)$-Bernstein condition with $B \leqslant \frac{1}{\Delta}$. However, the constant B may actually be much smaller than $\frac{1}{\Delta}$ in many situations where good experts are highly correlated. Hence, the parameter B is a more refined complexity parameter than Δ, which can be somewhat brittle.

We next consider the behavior of Decreasing Hedge under the Bernstein condition.

Proposition 5. *For every $T \geqslant 1$, there exists a $(1, 1)$-Bernstein stochastic instance on which the regret of Decreasing Hedge satisfies with $\eta_t = c_0\sqrt{(\log M)/t}$ satisfies $\mathbb{E}[R_T] \geqslant \frac{1}{3}\min(\frac{1}{c_0}\sqrt{T \log M}, T)$.*

In addition, for every i.i.d. (over time) stochastic instance with a unique best expert $i^* = \operatorname{argmin}_{1\leqslant i\leqslant M} \mathbb{E}[\ell_{i,t}]$, the regret Decreasing Hedge (with $c_0 \geqslant 1$) satisfies $\mathbb{E}[R_T] \geqslant \frac{1}{450 c_0^4 (\log M)^2 \Delta}$ for $T \geqslant \frac{1}{4\Delta^2}$, where $\Delta := \inf_{i\neq i^*} \mathbb{E}[\ell_{i,t} - \ell_{i^*,t}]$.

Proposition 5 is proved in Section 7.2.2. The first part of Proposition 5 states that Decreasing Hedge does not benefit from the Bernstein condition. This clarifies the advantage of tuning the learning rate in a data-dependent fashion (as second-order Hedge algorithms do), and in particular of using a larger learning rate on some stochastic instances, namely $(1, B)$-Bernstein instances with B small but with small sub-optimality gap Δ. Intuitively, the learning rate of Decreasing Hedge is large enough that it can rule out bad experts (with large enough gap Δ_i) at the optimal rate (*i.e.*, at time $(\log M)/\Delta_i^2$). However, once these bad experts are ruled out, the near-optimal experts (with small gap Δ_i) are ruled out late (after $(\log M)/\Delta_i^2$ rounds). On the other hand, since by the Bernstein assumption those experts are highly correlated with the best expert, the amount of noise on the relative losses of these near-optimal experts is small, so that a larger learning rate could be safely used and would enable to dismiss near-optimal experts sooner.

The second part of Proposition 5 completes this statement, by showing (together with the upper bound of Theorem 1) that the eventual regret of Decreasing Hedge on *any* stochastic instance is determined by the sub-optimality gap Δ, and scales (up to a $\log^3 M$ factor, depending on the number of near-optimal experts) as $\Theta(\frac{1}{\Delta})$. This characterizes the behavior of Decreasing Hedge on any stochastic instance.

5 Experiments

In this section, we illustrate our theoretical results by numerical experiments that compare the behavior of various Hedge algorithms in the stochastic regime.

Algorithms. We consider the following algorithms: `hedge` is Decreasing Hedge with the default learning rates $\eta_t = 2\sqrt{\log(M)/t}$, `hedge_constant` is Constant Hedge with constant learning rate $\eta_t = \sqrt{8\log(M)/T}$, `hedge_doubling` is Hedge with doubling trick with $c = \sqrt{8 \log M}$, `adahedge` is the AdaHedge algorithm from [9], which is a variant of the Hedge algorithm with a data-dependent tuning of the learning rate η_t (based on $\boldsymbol{\ell}_1, \dots, \boldsymbol{\ell}_{t-1}$). A related algorithm, namely Hedge with second-order tuning of the learning rate [7], performed similarly to AdaHedge on the examples considered below, and was therefore not included. `FTL` is Follow-the-Leader [6] which puts all mass on the expert with the smallest loss (breaking ties randomly). While FTL serves as a benchmark in the stochastic setting, unlike the other algorithms it lacks any guarantee in the adversarial regime, where its worst-case regret is *linear* in T.

Results. We report in Figure 1 the cumulative regrets of the considered algorithms in four examples.

(a) *Stochastic instance with a gap.* This is the standard instance considered in this paper. The losses are drawn independently from Bernoulli distribution (one of parameter 0.3, 2 of parameter 0.4 and 7 of parameter 0.5, so that $M = 10$ and $\Delta = 0.1$). The results of Figure 1a confirm our theoretical results: Decreasing Hedge achieves a small, constant regret which is close to that of AdaHedge and FTL, while Constant Hedge and Hedge with doubling trick suffer a larger regret of order $\sqrt{T}$ (note that, although the expected regret of Constant Hedge converges in this case, the value of this limit depends on its learning rate and hence on T).

(b) *"Hard" stochastic instance.* This example has a zero gap $\Delta = 0$ between the two leading experts and $M = 10$, which makes it "hard" from the standpoint of Theorem 1 (which no longer applies in this limit case). The losses are drawn from independent Bernoulli distributions, of parameters 0.5 for the 2 leading experts, and 0.7 for the 8 remaining ones. Although all algorithms suffer an unavoidable $\Theta(\sqrt{T})$ regret due to pure noise, Decreasing Hedge, AdaHedge and FTL achieve better regret than the two conservative Hedge variants (Figure 1b). This is due to the fact that for the former algorithms, the weights of suboptimal experts decrease quickly and only induce a constant regret.

(c) *Small loss for the best expert.* In this experiment, we illustrate one advantage of adaptive Hedge algorithms such as AdaHedge over Decreasing Hedge, namely the fact that they admit improved regret bounds when the leading expert has small regret. We considered in this experiment $M = 10$, $\Delta = 0.04$ and the leading expert is $\mathsf{Beta}(0.04, 0.96)$, then 4 $\mathsf{Beta}(0.08, 0.92)$, then 5 $\mathsf{Beta}(0.5, 0.5)$.

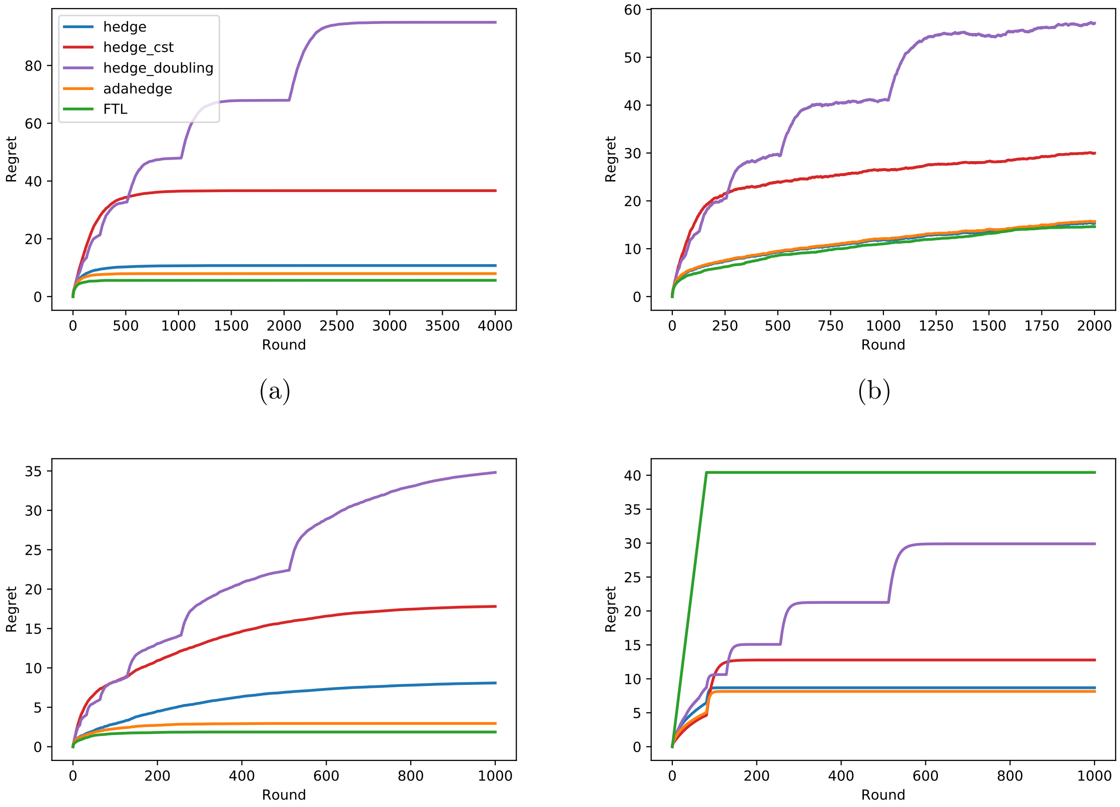
(d) *Adversarial with a gap instance.* This simple instance is not random, and satisfies the assumptions of Theorem 2. It is defined by $M = 3$, $\Delta = 0.04$, $\ell_{3,t} = \frac{3}{4}$ for $t \geqslant 1$, $(\ell_{1,t}, \ell_{2,t}) = (\frac{1}{2}, 0)$ if $t = 1$, $(0, 1)$ if $t \geqslant 80$ or if t is even, and $(1, 0)$ otherwise. FTL suffers linear regret in the first phase, while Constant Hedge and Hedge with doubling trick suffer $\Theta(\sqrt{T})$ during the second phase.

The results for the stochastic instances (a), (b) and (c) are averaged over 50 trials.

6 Conclusion

In this article, we carried the regret analysis of the standard exponential weights (Hedge) algorithm in the stochastic expert setting, closing a gap in the existing literature. Our analysis reveals a surprising phenomenon: despite being tuned for the worst-case adversarial setting and lacking any adaptive tuning of the learning rate, Decreasing Hedge achieves optimal regret in the stochastic setting. This property also enables to distinguish it qualitatively from other variants including the one with fixed (horizon-dependent) learning rate or the one with doubling trick, which both fail to adapt to gaps in the losses. To the best of our knowledge, this is the first result that shows the superiority of the decreasing learning rate over the doubling trick. In addition, it suggests that, even for a fixed time horizon T, the decreasing learning rate tuning should be favored over the constant one.

Finally, we showed that the regret of Decreasing Hedge on any stochastic instance is essentially characterized by the sub-optimality gap Δ. This shows that adaptive algorithms, including algorithms achieving second-order regret bounds, can actually outperform Decreasing Hedge on some stochastic instances that exhibit a more refined form of "easiness".



(a)

(b)

(c)

(d)

Figure 1: Cumulative regret of Hedge algorithms on four examples, see text for a precise description and discussion about the results. (a) Stochastic instance with a gap; (b) “Hard” stochastic instance; (c) Small loss for the best expert; (d) Adversarial with a gap instance.

A link with stochastic optimization. Our results have a similar flavor to a well-known result [20] about stochastic optimization: stochastic gradient descent (SGD) with learning rate $\eta_t \propto 1/\sqrt{t}$ (which is tuned for the convex case but not for the non-strongly convex case) and Polyak-Ruppert averaging achieves a fast $O(1/(\mu t))$ excess risk rate for μ-strongly convex problems, without the knowledge of μ. However, this link stops here since the two results are of a significantly different nature: the $O(1/(\mu t))$ rate is satisfied only by SGD with Polyak-Ruppert averaging, and it does not come from a regret bound; even in the μ-strongly convex case, it can be seen easily that SGD with step-size $\eta_t \propto 1/\sqrt{t}$ suffers a $\Theta(\sqrt{t})$ regret. In fact, the opposite phenomenon occurs: in stochastic optimization, SGD uses a *larger* $\Theta(1/\sqrt{t})$ step-size than the $\Theta(1/(\mu t))$ step size which exploits the knowledge of strong convexity, but the effect of this larger step-size is balanced by the averaging. By contrast, in the expert setting, Hedge uses a *smaller* $\Theta(\sqrt{(\log M)/t})$ learning rate than the constant, large enough learning rate which exploits the knowledge of the stochastic nature of the problem.

7 Proofs

We now provide the proofs of the results in the previous sections.

7.1 Upper bounds

In this section, we gather the proof of the regret upper bounds, namely Theorems 1 and 2 as well as Proposition 4.

7.1.1 Proof of Theorem 1

Let $t_0 = \left\lceil \frac{8 \log M}{\Delta^2} \right\rceil$, so that $\sqrt{t_0} \leqslant \sqrt{1 + \frac{8 \log M}{\Delta^2}} \leqslant 1 + \frac{\sqrt{8 \log M}}{\Delta}$ (since $\sqrt{a+b} \leqslant \sqrt{a} + \sqrt{b}$ for $a, b \geqslant 0$). The worst-case regret bound of Hedge (Proposition 1) shows that for $1 \leqslant T \leqslant t_0$:

$$R_T \leqslant \sqrt{T \log M} \leqslant \sqrt{t_0 \log M} \leqslant \sqrt{\log M} + \frac{2\sqrt{2} \log M}{\Delta} \leqslant \frac{4 \log M}{\Delta} \tag{18}$$

(since $\log M \geqslant 1$ as $M \geqslant 3$, $\Delta \leqslant 1$ and $2\sqrt{2} \leqslant 3$), which establishes (6) for $T \leqslant t_0$. In order to prove (6) for $T \geqslant t_0 + 1$, we start by decomposing the regret (with respect to i^*) as

$$R_{i^*,T} = L_T - L_{i^*,T} = L_{t_0} - L_{i^*,t_0} + \sum_{t=t_0+1}^{T} (\ell_t - \ell_{i^*,t}) \, . \tag{19}$$

Since $L_{t_0} - L_{i^*,t_0} \leqslant R_{t_0}$ is controlled by (18), it remains to upper bound the second term in (19). First, for every $t \geqslant t_0 + 1$,

$$\ell_t - \ell_{i^*,t} = \sum_{i \neq i^*} v_{i,t} (\ell_{i,t} - \ell_{i^*,t}) \, . \tag{20}$$

Since $\boldsymbol{\ell}_t$ is independent of $\boldsymbol{v}_t$ (which is $\mathcal{F}_{t-1} := \sigma(\boldsymbol{\ell}_1, \dots, \boldsymbol{\ell}_{t-1})$-measurable), taking the expectation in (20) yields

$$\mathbb{E}[\ell_t - \ell_{i^*,t}] = \sum_{i \neq i^*} \Delta_i \mathbb{E}[v_{i,t}] \, . \tag{21}$$

First, for every $i \neq i^*$, applying Hoeffding's inequality to the i.i.d. centered variables $Z_{i,t} := -\ell_{i,t} + \ell_{i^*,t} + \Delta_i$, which belong to $[-1+\Delta_i, 1+\Delta_i]$, yields

$$\begin{aligned} \mathbb{P}\left(L_{i,t-1} - L_{i^*,t-1} < \frac{\Delta_i(t-1)}{2}\right) &= \mathbb{P}\left(\sum_{s=1}^{t-1} Z_{i,s} > \frac{\Delta_i(t-1)}{2}\right) \\ &\leqslant e^{-\frac{t-1}{2}(\Delta_i/2)^2} \\ &= e^{-(t-1)\Delta_i^2/8}\,. \end{aligned} \tag{22}$$

On the other hand, if $L_{i,t-1} - L_{i^*,t-1} \geqslant \Delta_i(t-1)/2$, then

$$\begin{aligned} v_{i,t} &= \frac{e^{-\eta_t(L_{i,t-1}-L_{i^*,t-1})}}{1 + \sum_{j \neq i^*} e^{-\eta_t(L_{j,t-1}-L_{i^*,t-1})}} \\ &\leqslant e^{-2\sqrt{(\log M)/t} \times \Delta_i(t-1)/2} \\ &\leqslant e^{-\Delta_i\sqrt{(t-1)(\log M)/2}} \end{aligned} \tag{23}$$

since $t \leqslant 2(t-1)$. It follows from (23) and (22) that, for $t \geqslant t_0 + 1 \geqslant 2$,

$$\begin{aligned} \mathbb{E}[v_{i,t}] &\leqslant \mathbb{P}\left(L_{i,t-1} - L_{i^*,t-1} > \frac{\Delta_i(t-1)}{2}\right) + e^{-\Delta_i\sqrt{(t-1)(\log M)/2}} \\ &\leqslant e^{-(t-1)\Delta_i^2/8} + e^{-\Delta_i\sqrt{(t-1)(\log M)/2}}\,. \end{aligned} \tag{24}$$

Now, a simple analysis of functions shows that the functions $f_1(u) = ue^{-u}$ and $f_2(u) = ue^{-u^2/2}$ are decreasing on $[1, +\infty)$. Since $\Delta_i \geqslant \Delta$, this entails that

$$\Delta_i e^{-(t-1)\Delta_i^2/8} = \frac{2}{\sqrt{t-1}} f_2\left(\frac{\sqrt{t-1}\Delta_i}{2}\right) \leqslant \frac{2}{\sqrt{t-1}} f_2\left(\frac{\sqrt{t-1}\Delta}{2}\right) = \Delta e^{-(t-1)\Delta^2/8} \tag{25}$$

provided that $\frac{\sqrt{t-1}\Delta}{2} \geqslant 1$, *i.e.* $t \geqslant 1 + \frac{4}{\Delta^2}$, which is the case since $t \geqslant t_0 + 1 \geqslant 1 + \frac{16 \log M}{\Delta^2}$. Likewise,

$$\Delta_i e^{-\Delta_i\sqrt{(t-1)(\log M)/2}} \leqslant \Delta e^{-\Delta\sqrt{(t-1)(\log M)/2}} \tag{26}$$

if $\Delta\sqrt{(t-1)(\log M)/2} \geqslant 1$, *i.e.* $t \geqslant 1 + \frac{2}{(\log M)\Delta^2}$, which is ensured by $t \geqslant t_0 + 1$. It follows from (21), (24), (25) and (26) that for every $t \geqslant t_0 + 1$:

$$\begin{aligned} \mathbb{E}[\ell_t - \ell_{i^*,t}] &\leqslant M\Delta e^{-(t-1)\Delta^2/8} + M\Delta e^{-\Delta\sqrt{(t-1)(\log M)/2}} \\ &= \left(Me^{-t_0\Delta^2/8}\right)\left(\Delta e^{-(t-t_0-1)\Delta^2/8}\right) + \left(Me^{-\Delta\sqrt{(t-1)(\log M)/8}}\right)\left(\Delta e^{-\Delta\sqrt{(t-1)(\log M)/8}}\right) \\ &\leqslant \Delta e^{-(t-t_0-1)\Delta^2/8} + \Delta e^{-\Delta\sqrt{(t-1)/8}} \end{aligned} \tag{27}$$

where inequality (27) comes from the bound $Me^{-t_0\Delta^2/8} \leqslant 1$ (since $t_0 \geqslant \frac{8 \log M}{\Delta^2}$) and from the fact that $Me^{-\Delta\sqrt{(t-1)(\log M)/8}} \leqslant 1$ amounts to $t \geqslant 1 + \frac{8 \log M}{\Delta^2}$, that is, to $t \geqslant t_0 + 1$. Summing inequality (27) yields, for every $T \geqslant t_0+1$,

$$\begin{aligned}
\mathbb{E}\left[\sum_{t=t_0+1}^{T}(\ell_t-\ell_{i^*,t})\right] &\leqslant \sum_{t=t_0+1}^{T}\left\{\Delta e^{-(t-t_0-1)\Delta^2/8}+\Delta e^{-\Delta\sqrt{(t-1)/8}}\right\}\\
&\leqslant \Delta\sum_{t\geqslant 0}e^{-t\Delta^2/8}+\Delta\sum_{t\geqslant 1}e^{-(\Delta/\sqrt{8})\sqrt{t}}\\
&\leqslant \Delta\left(1+\frac{8}{\Delta^2}\right)+\Delta\times\frac{2}{(\Delta/\sqrt{8})^2} \qquad (28)\\
&\leqslant \frac{25}{\Delta} \qquad (29)
\end{aligned}$$

where inequality (28) comes from Lemma 1 below.

Finally, combining inequalities (18) and (28) yields the expected pseudo-regret bound $\mathbb{E}[R_{i^*,T}] \leqslant \frac{4\log M+25}{\Delta}$. In order to obtain the expected regret bound (6), it remains to note that

$$R_T = R_{i^*,T} + \left(L_{i^*,T} - \min_{1\leqslant i\leqslant M} L_{i,T}\right)$$

and use the fact that $\mathbb{E}[L_{i^*,T}-\min_{1\leqslant i\leqslant M} L_{i,T}] \leqslant \frac{2}{\Delta}$ for $T \geqslant t_0+1 \geqslant \frac{4\log M}{\Delta^2}$, by Lemma 2 below. □

Lemma 1. *For every $\alpha > 0$,*

$$\sum_{t\geqslant 1} e^{-\alpha t} \leqslant \frac{1}{\alpha} \qquad (30)$$

$$\sum_{t\geqslant 1} e^{-\alpha\sqrt{t}} \leqslant \frac{2}{\alpha^2}\,. \qquad (31)$$

Proof of Lemma 1. Since the functions $t \mapsto e^{-\alpha t}$ and $t \mapsto e^{-\alpha\sqrt{t}}$ are decreasing on $\mathbf{R}^+$, we have

$$\sum_{t\geqslant 1} e^{-\alpha t} \leqslant \int_0^{\infty} e^{-\alpha t}\mathrm{d}t = \frac{1}{\alpha}\,;$$

$$\sum_{t\geqslant 1} e^{-\alpha\sqrt{t}} \leqslant \int_0^{+\infty} e^{-\alpha\sqrt{t}}\mathrm{d}t \underset{u=\alpha\sqrt{t}}{=} \frac{2}{\alpha^2}\int_0^{+\infty} u e^{-u}\mathrm{d}u = \frac{2}{\alpha^2}\,.$$

□

Lemma 2. *Under the assumptions of Theorem 1, for every $T \geqslant \frac{4\log M}{\Delta^2}$, we have*

$$\mathbb{E}\left[L_{i^*,T} - \min_{1\leqslant i\leqslant T} L_{i,T}\right] \leqslant \frac{1.1}{\Delta}\,. \qquad (32)$$

Proof of Lemma 2. For every $a \geqslant 0$, Hoeffding's inequality (applied to the i.i.d. centered variables $\ell_{i^*,t} - \ell_{i,t} + \Delta_i \in [-1+\Delta_i, 1+\Delta_i]$, $1 \leqslant t \leqslant T$) entails

$$\begin{aligned}
\mathbb{P}\left(L_{i^*,T} - \min_{1\leqslant i\leqslant T} L_{i,T} \geqslant a\right) &\leqslant \sum_{i\neq i^*}\mathbb{P}\left(L_{i^*,T} - L_{i,T} + \Delta_i T \geqslant \Delta_i T + a\right)\\
&\leqslant \sum_{i\neq i^*} e^{-(\Delta_i T+a)^2/(2T)} \qquad (33)\\
&\leqslant M e^{-T\Delta^2/2} e^{-a^2/(2T)}\\
&\leqslant e^{-T\Delta^2/4} e^{-a^2/(2T)}\,, \qquad (34)
\end{aligned}$$

where inequality (34) comes from the fact that $Me^{-T\Delta^2/4} \leqslant 1$ since $T \geqslant \frac{4\log M}{\Delta^2}$. Since the random variable $L_{i^*,T} - \min_{1\leqslant i\leqslant T} L_{i,T}$ is nonnegative, this implies that

$$\begin{aligned}
\mathbb{E}\left[L_{i^*,T} - \min_{1\leqslant i\leqslant T} L_{i,T}\right] &= \int_0^\infty \mathbb{P}\left(L_{i^*,T} - \min_{1\leqslant i\leqslant T} L_{i,T} \geqslant a\right) \mathrm{d}a \\
&\leqslant e^{-T\Delta^2/4} \int_0^\infty e^{-a^2/(2T)} \mathrm{d}a \\
&= \sqrt{\frac{\pi}{2}} \cdot \sqrt{T} e^{-T\Delta^2/4} \\
&= \frac{\sqrt{\pi}}{\Delta} \left[\Delta\sqrt{T/2} \cdot e^{-(\Delta\sqrt{T/2})^2/2}\right] \\
&\leqslant \frac{\sqrt{\pi/e}}{\Delta} \qquad (35)
\end{aligned}$$

where inequality (35) comes from the fact that the function $u \mapsto ue^{-u^2/2}$ attains its maximum on $\mathbf{R}^+$ at $u = 1$. This concludes the proof, since $\sqrt{\pi/e} \leqslant 1.1$. □

7.1.2 Proof of Theorem 2 and Corollary 1

Let t_0 be the smallest integer $t \geqslant 1$ such that $Me^{-c_0\Delta\sqrt{t\log(M)/8}} \leqslant \Delta$, namely $t_0 = \left\lceil \frac{8}{c_0^2\Delta^2} \frac{\log^2(M/\Delta)}{\log M} \right\rceil$. Note that $\sqrt{t_0} \leqslant \sqrt{1 + \frac{8}{c_0^2\Delta^2} \frac{\log^2(M/\Delta)}{\log M}} \leqslant 1 + \frac{\sqrt{8}}{c_0\Delta} \frac{\log(M/\Delta)}{\sqrt{\log M}}$. Let $t_1 := t_0 \vee \tau_0$. For every $T \leqslant t_1$, the regret bound in the assumption of Theorem 2 implies

$$\begin{aligned}
R_T &\leqslant c_1\sqrt{T\log M} \\
&\leqslant c_1\sqrt{\tau_0 \log M} + c_1\sqrt{t_0\log M} \\
&\leqslant c_1\sqrt{\tau_0 \log M} + c_1\sqrt{\log M} + \frac{\sqrt{8}\log(M/\Delta)}{c_0\Delta} \qquad (36)
\end{aligned}$$

which implies (9) with $c_2 = c_1 + \frac{\sqrt{8}}{c_0}$ and $c_3 = \frac{\sqrt{8}}{c_0}$ (since $\sqrt{\log M} \leqslant \frac{\log M}{\Delta}$). From now on, assume that $T \geqslant t_1 + 1$. Since $T \geqslant \tau_0$, we have $R_T = L_T - L_{i^*,T}$, so that

$$R_T = L_{t_1} - L_{i^*,t_1} + \sum_{t=t_1+1}^{T} (\ell_t - \ell_{i^*,t}) \ . \qquad (37)$$

In addition, we have for $t \geqslant t_1 + 1$

$$
\begin{aligned}
\ell_t - \ell_{i^*,t} &= \sum_{i \neq i^*} v_{i,t}(\ell_{i,t} - \ell_{i^*,t}) \\
&\leqslant \sum_{i \neq i^*} v_{i,t} \\
&= \sum_{i \neq i^*} \frac{e^{-\eta_t(L_{i,t-1} - L_{i^*,t-1})}}{1 + \sum_{j \neq i^*} e^{-\eta_t(L_{j,t-1} - L_{i^*,t-1})}} \\
&\leqslant \sum_{i \neq i^*} e^{-c_0 \sqrt{(\log M)/t} \times \Delta(t-1)} \qquad (38) \\
&\leqslant M e^{-c_0 \Delta \sqrt{(t-1)(\log M)/2}} \\
&\leqslant \left(M e^{-c_0 \Delta \sqrt{t_0 (\log M)/8}}\right) e^{-c_0 \Delta \sqrt{(t-1)/8}} \qquad (39) \\
&\leqslant \Delta e^{-c_0 \Delta \sqrt{(t-1)/8}} \qquad (40)
\end{aligned}
$$

where (38) comes from the fact that $\eta_t \geqslant c_0 \sqrt{(\log M)/t}$ and $L_{i,t-1} - L_{i^*,t-1} \geqslant \Delta(t-1)$ (since $t - 1 \geqslant t_1 \geqslant \tau_0$), (39) from the fact that $t - 1 \geqslant t_0$ and $\log M \geqslant 1$, and (40) from the fact that $M e^{-c_0 \Delta \sqrt{t_0 (\log M)/8}} \leqslant \Delta$. Summing inequality (40), we obtain

$$
\begin{aligned}
\sum_{t=t_1+1}^{T} (\ell_t - \ell_{i^*,t}) &\leqslant \sum_{t=t_1+1}^{T} \Delta e^{-c_0 \Delta \sqrt{(t-1)/8}} \\
&\leqslant \Delta \sum_{t \geqslant 1} e^{-c_0 \Delta \sqrt{t/8}} \\
&\leqslant \Delta \times \frac{2}{(c_0 \Delta / \sqrt{8})^2} \qquad (41) \\
&= \frac{16}{c_0^2 \Delta} \qquad (42)
\end{aligned}
$$

where (41) follows from Lemma 1. Combining (37), (36) and (42) proves Theorem 2 with $c_2 = c_1 + \frac{\sqrt{8}}{c_0}$, $c_3 = \frac{\sqrt{8}}{c_0}$ and $c_4 = \frac{16}{c_0^2 \Delta}$.

Proof of Corollary 1. Define $\tau = \sup\{t \geqslant 0, \exists i \neq i^*, L_{i,t} - L_{i^*,t} \leqslant \frac{\Delta t}{2}\}$. By Lemma 3 below, for every $\varepsilon > 0$ we have, with probability at least $1 - \varepsilon$, $\tau \leqslant 8(\log M + \log \varepsilon^{-1})/\Delta^2$. By Theorem 2, this implies that, with probability at least $1 - \varepsilon$,

$$
\begin{aligned}
R_T &\leqslant c_1 \sqrt{\tau \log M} + \frac{c_2 \log M + c_3 \log \Delta^{-1} + c_4}{\Delta/2} \\
&\leqslant \left(c_1 \sqrt{8} + 2c_2\right) \frac{\log M}{\Delta} + c_1 \frac{\sqrt{8 \log M \log \varepsilon^{-1}}}{\Delta} + 2c_3 \frac{\log \Delta^{-1}}{\Delta} + \frac{2c_4}{\Delta}
\end{aligned}
$$

where c_2, c_3, c_4 are the constants of Theorem 2. The bound (11) on the expected regret is obtained similarly from Theorem 2, by using the fact that

$$
\mathbb{E}[\sqrt{\tau \log M}] \leqslant \sqrt{\mathbb{E}[\tau] \log M} \leqslant \sqrt{\log M} \sqrt{1 + \frac{8(\log M + 1)}{\Delta^2}} \leqslant \sqrt{\log M} \left(1 + \frac{\sqrt{8 \log M} + 1}{\Delta}\right)
$$

which is smaller than $(2+\sqrt{8})(\log M)/\Delta \leqslant 5(\log M)/\Delta$ since $M \geqslant 3$ and $\Delta \leqslant 1$. □

Lemma 3. *Let $(\ell_{i,t})_{1\leqslant i\leqslant M, t\geqslant 1}$ be as in Theorem 1. Denote $\tau = \sup\{t \geqslant 0, \exists i \neq i^*, L_{i,t} - L_{i^*,t} \leqslant \frac{\Delta t}{2}\}$. We have*

$$\mathbb{E}[\tau] \leqslant 1 + \frac{8(\log M + 1)}{\Delta^2}\,, \tag{43}$$

and for every $\varepsilon \in (0,1)$,

$$\mathbb{P}\Big(\tau \geqslant \frac{8(\log M + \log \varepsilon^{-1})}{\Delta^2}\Big) \leqslant \varepsilon\,. \tag{44}$$

Proof of Lemma 3. For every $i \neq i^*$ and $t \geqslant 1$, let $\Delta_{i,t} := \mathbb{E}[\ell_{i,t} - \ell_{i^*,t} \,|\, \mathcal{F}_{t-1}]$. Hoeffding-Azuma's maximal inequality applied to the $(\mathcal{F}_t)_{t\geqslant 1}$-martingale difference sequence $Z_{i,t} = -(L_{i,t} - L_{i^*,t}) + \Delta_{i,t}$ (such that $\Delta_{i,t} - 1 \leqslant Z_{i,t} \leqslant \Delta_{i,t} + 1$), combined with the fact that $\Delta_{i,t} \geqslant \Delta$, implies that

$$\mathbb{P}\left(\exists t \geqslant t_0, L_{i,t} - L_{i^*,t} \leqslant \frac{\Delta t}{2}\right) \leqslant \mathbb{P}\left(\sup_{t\geqslant t_0} \frac{1}{t}\left(\sum_{s=1}^{t} Z_{i,s}\right) \geqslant \frac{\Delta}{2}\right) \leqslant e^{-t_0\Delta^2/8}\,. \tag{45}$$

By a union bound, equation (45) implies that

$$\mathbb{P}\left(\tau \geqslant t_0\right) \leqslant M e^{-t_0\Delta^2/8}\,. \tag{46}$$

Solving for the probability level in (46) yields the high probability bound (44) on τ. The bound on τ in expectation (43) ensues by integrating the high-probability bound over ε. □

We recall Hoeffding-Azuma's maximal inequality for bounded martingale difference sequences [12, 2].

Proposition 6 (Hoeffding-Azuma's maximal inequality)**.** *Let $(Z_t)_{t\geqslant 1}$ be a sequence of random variables adapted to a filtration $(\mathcal{F}_t)_{t\geqslant 1}$. Assume that Z_t is a martingale difference sequence: $\mathbb{E}[Z_t \,|\, \mathcal{F}_{t-1}] = 0$ for any $t \geqslant 1$, and that $A_t - 1 \leqslant Z_t \leqslant A_t + 1$ almost surely, where A_t is $\mathcal{F}_{t-1}$-measurable. Then, denoting $S_n := \sum_{t=1}^n Z_t$, we have for every $n \geqslant 1$ and $a \geqslant 0$:*

$$\mathbb{P}\left(\sup_{m\geqslant n} \frac{S_m}{m} \geqslant a\right) \leqslant e^{-na^2/2}\,. \tag{47}$$

7.1.3 Proof of Proposition 4

By convexity of $x \mapsto x^2$ and concavity of $x \mapsto x^\beta$, we have:

$$\begin{aligned}
\mathbb{E}[(\ell_t - \ell_{i^*,t})^2] &\leqslant \mathbb{E}\left[\sum_{i=1}^M v_{i,t}(\ell_{i,t} - \ell_{i^*,t})^2\right] && (48)\\
&= \mathbb{E}\left[\sum_{i=1}^M v_{i,t}\mathbb{E}\left[(\ell_{i,t} - \ell_{i^*,t})^2 \,|\, \mathcal{F}_{t-1}\right]\right] \\
&\leqslant B\mathbb{E}\left[\sum_{i=1}^M v_{i,t}\mathbb{E}\left[\ell_{i,t} - \ell_{i^*,t} \,|\, \mathcal{F}_{t-1}\right]^\beta\right] && (49)\\
&\leqslant B\mathbb{E}\left[\sum_{i=1}^M v_{i,t}\mathbb{E}\left[\ell_{i,t} - \ell_{i^*,t} \,|\, \mathcal{F}_{t-1}\right]\right]^\beta && (50)\\
&= B\mathbb{E}[\ell_t - \ell_{i^*,t}]^\beta && (51)
\end{aligned}$$

where inequalities (48) and (50) come from Jensen's inequality, and (49) from the Bernstein condition (15). Taking the expectation of the regret bound (16), we obtain

$$\begin{aligned}
\mathbb{E}[R_{i^*,T}] &\leqslant \mathbb{E}\left[C_1\sqrt{(\log M)\sum_{t=1}^T(\ell_t-\ell_{i^*,t})^2}+C_2\log M\right]\\
&\leqslant C_1\sqrt{(\log M)\sum_{t=1}^T\mathbb{E}\left[(\ell_t-\ell_{i^*,t})^2\right]}+C_2\log M \qquad (52)\\
&\leqslant C_1\sqrt{(\log M)B\sum_{t=1}^T\mathbb{E}\left[\ell_t-\ell_{i^*,t}\right]^\beta}+C_2\log M\\
&= C_1\sqrt{BT\log M}\left(\frac{1}{T}\sum_{t=1}^T\mathbb{E}\left[\ell_t-\ell_{i^*,t}\right]^\beta\right)^{1/2}+C_2\log M\\
&\leqslant C_1\sqrt{BT\log M}\left(\frac{\mathbb{E}[R_{i^*,T}]}{T}\right)^{\beta/2}+C_2\log M \qquad (53)
\end{aligned}$$

where inequalities (52) and (53) come from Jensen's inequality. Letting $r=\frac{\mathbb{E}[R_{i^*,T}]}{T}$ and $u=\frac{\log M}{T}$, inequality (53) writes $r\leqslant C_1\sqrt{Bu}r^{\beta/2}+C_2u$. This implies that (depending on which of these two terms is larger) either $r\leqslant 2C_2u$, or $r\leqslant 2C_1\sqrt{Bu}r^{\beta/2}$, and the latter condition amounts to $r\leqslant(2C_1)^{2/(2-\beta)}(Bu)^{1/(2-\beta)}$. This entails that

$$r\leqslant(2C_1)^{\frac{2}{2-\beta}}(Bu)^{\frac{1}{2-\beta}}+2C_2u\,,$$

which amounts to

$$\mathbb{E}[R_{i^*,T}]\leqslant C_3(B\log M)^{\frac{1}{2-\beta}}T^{\frac{1-\beta}{2-\beta}}+C_4\log M \qquad (54)$$

where $C_3=(2C_1)^{2/(2-\beta)}\leqslant\max(1,4C_1^2)$ and $C_4=2C_2$.

7.2 Lower bounds

We now provide the proofs of the regret lower bounds, namely Propositions 3, 5 and 2.

7.2.1 Proof of Proposition 3

Denoting $(v_{i,t})_{1\leqslant i\leqslant M}$ the weights selected by the Constant Hedge algorithm at time t, and letting $c=c_0\sqrt{\log M}$, we have

$$\begin{aligned}
R_T &= \sum_{t=1}^T\sum_{i=2}^M v_{i,t}(\ell_{i,t}-\ell_{1,t})\\
&= \sum_{t=1}^T\sum_{i=2}^M\frac{\exp\left(-\frac{c}{\sqrt{T}}(L_{i,t-1}-L_{1,t-1})\right)}{1+\sum_{2\leqslant i'\leqslant M}\exp\left(-\frac{c}{\sqrt{T}}(L_{i',t-1}-L_{1,t-1})\right)}\\
&= \sum_{t=1}^T\frac{(M-1)\exp\left(-\frac{c}{\sqrt{T}}(t-1)\right)}{1+(M-1)\exp\left(-\frac{c}{\sqrt{T}}(t-1)\right)}\,. \qquad (55)
\end{aligned}$$

Now, let $t_0 \geqslant 0$ be the largest integer such that $(M-1)\exp(-\frac{c}{\sqrt{T}}t) \geqslant 1/2$, namely $t_0 = \left\lfloor \frac{\sqrt{T}}{c}\log(2(M-1)) \right\rfloor$. It follows from Equation (55) that

$$R_T \geqslant \sum_{t=1}^{T\wedge(t_0+1)} \frac{(M-1)\exp\big(-\frac{c}{\sqrt{T}}(t-1)\big)}{1+(M-1)\exp\big(-\frac{c}{\sqrt{T}}(t-1)\big)} \geqslant \frac{1}{3}\min(T, t_0+1) \tag{56}$$

where the second inequality comes from the fact that $\frac{x}{1+x} \geqslant \frac{1}{3}$ for $x \geqslant \frac{1}{2}$, which we apply to $x = (M-1)\exp(-\frac{c}{\sqrt{T}}(t-1)) \geqslant \frac{1}{2}$ for $t \leqslant T \wedge (t_0+1) \leqslant t_0+1$. In order to establish inequality (13), it remains to note that

$$t_0 + 1 \geqslant \frac{\sqrt{T}}{c}\log\big(2(M-1)\big) \geqslant \frac{\sqrt{T\log M}}{c_0}\,,$$

since $2(M-1) \geqslant M$ and $c = \sqrt{c_0 \log M}$.

Now, consider the Hedge algorithm with doubling trick. Assume that $T \geqslant 2$, and let $k \geqslant 1$ such that $T_k \leqslant T < T_{k+1}$. Since $R_T = \sum_{t=1}^{T}\sum_{2\leqslant i\leqslant M} v_{i,t}(\ell_{i,t} - \ell_{1,t})$ and each of the terms in the sum is nonnegative, R_T is lower bounded by the cumulative regret on the period $[\![T_{k-1}, T_k - 1]\!]$. During this period of length T_{k-1}, the algorithm reduces to the Hedge algorithm with constant learning rate $c_0\sqrt{\log(M)/T_{k-1}}$, so that the above bound (13) applies; further bounding $T_{k-1} \geqslant \frac{T}{4}$ establishes (14).

7.2.2 Proof of Proposition 5

We start by establishing the first part of Proposition 5. Consider the constant losses $\ell_{1,t} = 0$, $\ell_{i,t} = \Delta$ where $\Delta = 1 \wedge c_0^{-1}\sqrt{(\log M)/T}$. These losses satisfy the $(1,1)$-Bernstein condition since, for every $i > 1$, $\mathbb{E}[(\ell_{i,t} - \ell_{1,t})^2] = \Delta^2 \leqslant \Delta = \mathbb{E}[\ell_{i,t} - \ell_{1,t}]$. On the other hand, the regret of the Hedge algorithm with learning rate $\eta_t = c_0\sqrt{(\log M)/t}$ writes

$$\begin{aligned}
R_T &= \sum_{t=1}^{T}\sum_{i\neq 1} v_{i,t}(\ell_{i,t} - \ell_{1,t}) \\
&= \Delta \sum_{t=1}^{T} \frac{(M-1)e^{-\eta_t\Delta(t-1)}}{1+(M-1)e^{-\eta_t\Delta(t-1)}} \\
&\geqslant \frac{\Delta}{3}\sum_{t=1}^{T} \mathbf{1}\left((M-1)e^{-\eta_t\Delta(t-1)} \geqslant \frac{1}{2}\right) \\
&\geqslant \frac{\Delta}{3}\sum_{t=1}^{T} \mathbf{1}\left(Me^{-c_0\Delta\sqrt{(t-1)\log M}} \geqslant 1\right) \qquad (57) \\
&\geqslant \frac{\Delta}{3} \times \min\left(\frac{\log M}{c_0^2\Delta^2}, T\right) \\
&= \frac{1}{3}\min\left(\frac{1}{c_0}\sqrt{T\log M}, T\right)\,, \qquad (58)
\end{aligned}$$

where (57) relies on the inequalities $2(M-1) \geqslant M$ and $(t-1)/\sqrt{t} \geqslant \sqrt{t-1}$ for $M \geqslant 2, t \geqslant 1$, while (58) is obtained by noting that $(\log M)/(c_0^2\Delta^2) \geqslant T$ since $\Delta \leqslant c_0^{-1}\sqrt{(\log M)/T}$ and substituting for Δ.

We now establish the second part of Proposition 5. Assume that the loss vectors $\boldsymbol{\ell}_1, \boldsymbol{\ell}_2, \dots$ are i.i.d., and denote $i^* = \operatorname{argmin}_{1 \leqslant i \leqslant M} \mathbb{E}[\ell_{i,t}]$ (which is assumed to be unique), $\Delta = \min_{i \neq i^*} \Delta_i > 0$ where $\Delta_i = \mathbb{E}[\ell_{i,t} - \ell_{i^*,t}]$ and $j \in \{1, \dots, M\}$ such that $\Delta_j = \Delta$. The Decreasing Hedge algorithm with learning rate $\eta_t = c_0 \sqrt{(\log M)/t}$ satisfies

$$\begin{aligned}
\mathbb{E}[R_T] &\geqslant \mathbb{E}[L_T - L_{i^*,T}] \\
&= \sum_{t=1}^{T} \sum_{i \neq i^*} \mathbb{E}[v_{i,t}] \Delta_i \\
&\geqslant \Delta \sum_{t=1}^{T} \mathbb{E}\left[\frac{\sum_{i \neq i^*} e^{-\eta_t (L_{i,t-1} - L_{i^*,t-1})}}{1 + \sum_{i \neq i^*} e^{-\eta_t (L_{i,t-1} - L_{i^*,t-1})}} \right] \\
&\geqslant \Delta \sum_{t=1}^{T} \mathbb{E}\left[\frac{e^{-\eta_t (L_{j,t-1} - L_{i^*,t-1})}}{1 + e^{-\eta_t (L_{j,t-1} - L_{i^*,t-1})}} \right] \qquad (59) \\
&\geqslant \frac{\Delta}{3} \sum_{t=1}^{T} \mathbb{E}\left[\mathbf{1}\left(e^{-\eta_t (L_{j,t-1} - L_{i^*,t-1})} \geqslant \frac{1}{2} \right) \right] \\
&= \frac{\Delta}{3} \sum_{t=1}^{T} \mathbb{P}\left(\eta_t (L_{j,t-1} - L_{i^*,t-1}) \leqslant \log 2 \right) \qquad (60)
\end{aligned}$$

where (59) relies on the fact that the function $x \mapsto \frac{x}{1+x}$ is increasing on $\mathbf{R}^+$. Denoting $a = (\log 2)/(c_0 \sqrt{\log M})$, we have for every $1 \leqslant t \leqslant 1 + \frac{a^2}{4\Delta^2}$:

$$\begin{aligned}
\mathbb{P}\left(\eta_t (L_{j,t-1} - L_{i^*,t-1}) > \log 2 \right) &= \mathbb{P}\left(L_{j,t-1} - L_{i^*,t-1} - \Delta(t-1) > a\sqrt{t} - \Delta(t-1) \right) \\
&\leqslant \mathbb{P}\left(L_{j,t-1} - L_{i^*,t-1} - \Delta(t-1) > \frac{a\sqrt{t-1}}{2} \right) \qquad (61) \\
&\leqslant e^{-a^2/8} \qquad (62)
\end{aligned}$$

where inequality (61) stems from the fact that $\Delta(t-1) \leqslant \frac{a\sqrt{t-1}}{2}$ (since $t \leqslant 1 + \frac{a^2}{4\Delta^2}$), while (62) is a consequence of Hoeffding's bound applied to the i.i.d. $[-1-\Delta, 1-\Delta]$-valued random variables $\ell_{j,s} - \ell_{i^*,s} - \Delta$, $1 \leqslant s \leqslant t-1$. Assuming that $c_0 \geqslant 1$, we have $a \leqslant \sqrt{\log 2} \leqslant 1$, so that by concavity of the function $x \mapsto 1 - e^{-x/8}$, $1 - e^{-a^2/8} \geqslant (1 - e^{-1/8}) a^2$. Combining this with inequalities (60) and (62) and using the fact that $\left\lfloor 1 + \frac{a^2}{4\Delta^2} \right\rfloor \geqslant \frac{a^2}{4\Delta^2}$, we obtain for $T \geqslant \frac{1}{4\Delta^2} \geqslant \frac{a^2}{4\Delta^2}$:

$$\mathbb{E}\left[R_T \right] \geqslant \frac{\Delta}{3} \min\left(\frac{a^2}{4\Delta^2}, T \right) (1 - e^{-1/8}) a^2 = \frac{(1 - e^{-1/8}) a^4}{12\Delta} \geqslant \frac{1}{450 c_0^4 (\log M)^2 \Delta}, \qquad (63)$$

where the last inequality comes from the fact that $(\log 2)^4 (1 - e^{-1/8})/12 \geqslant \frac{1}{450}$.

7.2.3 Proof of Proposition 2

Lower bound through cumulative Bayes risk. Fix M, Δ and T as in Proposition 2. For $i^* \in \{1, \dots, M\}$, denote $\mathbb{P}_{i^*}$ the following distribution on $[0,1]^{M \times T}$: if $(\ell_{i,t})_{1 \leqslant i \leqslant M, 1 \leqslant t \leqslant T} \sim \mathbb{P}_{i^*}$, then the variables $\ell_{i,t}$ are independent Bernoulli variables, of parameter $\frac{1}{2}$ if $i = i^*$ and $\frac{1}{2} + \Delta$ otherwise. Let $\mathcal{A} = (A_t)_{1 \leqslant t \leqslant T}$

be any Hedging algorithm, where $A_t : [0,1]^{M\times(t-1)} \to \mathcal{P}_M$ maps past losses $(\boldsymbol{\ell}_1, \dots, \boldsymbol{\ell}_{t-1})$ to an element of the probability simplex $\mathcal{P}_M \subset \mathbf{R}^M$ on $\{1,\dots,M\}$. Then, denoting $\mathbb{E}_{i^*}$ the expectation under $\mathbb{P}_{i^*}$ for $i^* \in \{1,\dots,M\}$, and $R_T = R_T^{\mathcal{A}}$ the regret of algorithm $\mathcal{A}$, we have

$$\sup_{1\leqslant i^*\leqslant M} \mathbb{E}_{i^*}[R_T] \geqslant \frac{1}{M}\sum_{i^*=1}^M \mathbb{E}_{i^*}[R_T] \geqslant \frac{1}{M}\sum_{i^*=1}^M \mathbb{E}_{i^*}[L_T - L_{i^*,T}]\,. \tag{64}$$

In addition, since for every i^*, $i \neq i^*$ and $1 \leqslant t \leqslant T$, $\boldsymbol{\ell}_t$ is independent of $\boldsymbol{\ell}_1, \dots, \boldsymbol{\ell}_{t-1}$ and hence of $\boldsymbol{v}_t := A_t(\boldsymbol{\ell}_1, \dots, \boldsymbol{\ell}_{t-1})$ under $\mathbb{P}_{i^*}$, we have $\mathbb{E}_{i^*}[v_{i,t}(\ell_{i,t} - \ell_{i^*,t})] = \Delta \mathbb{E}_{i^*}[v_{i,t}]$. Therefore,

$$\mathbb{E}_{i^*}[L_T - L_{i^*,T}] = \sum_{t=1}^T \sum_{i\neq i^*} \mathbb{E}_{i^*}[v_{i,t}(\ell_{i,t} - \ell_{i^*,t})] = \Delta \sum_{t=1}^T \sum_{i\neq i^*} \mathbb{E}_{i^*}[v_{i,t}] = \Delta \sum_{t=1}^T \mathbb{E}_{i^*}[1 - v_{i^*,t}]\,,$$

so that inequality (64) becomes:

$$\sup_{1\leqslant i^*\leqslant M} \mathbb{E}_{i^*}[R_T] \geqslant \Delta \sum_{t=1}^T \frac{1}{M}\sum_{i^*=1}^M \mathbb{E}_{i^*}[1 - v_{i^*,t}]\,. \tag{65}$$

For $0 \leqslant t \leqslant T-1$, we will provide a lower bound on $\frac{1}{M}\sum_{i^*=1}^M \mathbb{E}_{i^*}[1 - v_{i^*,t+1}]$ by interpreting it as a Bayes risk. Consider the decision problem with space of observations $[0,1]^{M\times t}$, statistical model $\Theta = \{\mathbb{P}_{i^*}, 1 \leqslant i^* \leqslant M\}$, state of actions $\mathcal{P}_M$ and loss function $\mathcal{L}(i^*, \boldsymbol{v}) = 1 - v_{i^*}$ for $i^* \in \{1,\dots,M\}$ and $\boldsymbol{v} = (v_i)_{1\leqslant i\leqslant M} \in \mathcal{P}_M$. The risk of a decision rule $A : [0,1]^{M\times t} \to \mathcal{P}_M$ on instance i^* is then

$$\mathcal{R}_t(i^*, A) = \mathbb{E}_{i^*}[\mathcal{L}(i^*, A(\boldsymbol{\ell}_1, \dots, \boldsymbol{\ell}_t))]\,.$$

Finally, let π be the uniform distribution on $\{1,\dots,M\}$. The Bayes risk of A under π is then $\mathcal{R}_t(\pi, A) := \mathbb{E}_{i^*\sim\pi}[\mathcal{L}(i^*, A)]$, so that:

$$\frac{1}{M}\sum_{i^*=1}^M \mathbb{E}_{i^*}[1 - v_{i^*,t+1}] = \mathcal{R}_t(\pi, A_{t+1}) \geqslant \inf_A \mathcal{R}_t(\pi, A)\,. \tag{66}$$

Bayes risk and Bayes decision rule. We will now control the Bayes risk $\inf_A \mathcal{R}_t(\pi, A)$. The first step is to determine the Bayes rule under π. First, the posterior distribution $\pi(\cdot \,|\, \boldsymbol{\ell}_1, \dots, \boldsymbol{\ell}_t)$ on Θ satisfies, noting that $L_{i,t} = |\{1 \leqslant s \leqslant t : \ell_{i,s} = 1\}|$ since the losses are binary,

$$\begin{aligned}
\pi(i^* \,|\, \boldsymbol{\ell}_1, \dots, \boldsymbol{\ell}_t) &\propto \mathbb{P}_{i^*}(\boldsymbol{\ell}_1, \dots, \boldsymbol{\ell}_t) \\
&= \frac{1}{2^t} \prod_{i\neq i^*} \Big(\frac{1}{2} + \Delta\Big)^{L_{i,t}} \Big(\frac{1}{2} - \Delta\Big)^{t-L_{i,t}} \\
&= \Big(\frac{1}{2} + \Delta\Big)^{-L_{i^*,t}} \Big(\frac{1}{2} - \Delta\Big)^{-(t-L_{i^*,t})} \times \frac{1}{2^t} \prod_{i=1}^M \Big(\frac{1}{2} + \Delta\Big)^{L_{i,t}} \Big(\frac{1}{2} - \Delta\Big)^{t-L_{i,t}} \\
&\propto \left(\frac{1-2\Delta}{1+2\Delta}\right)^{2L_{i^*,t}}\,.
\end{aligned} \tag{67}$$

In addition, note that for every distribution ρ over Θ, the average loss of the action $\boldsymbol{v} \in \mathcal{P}_M$ is

$$\mathbb{E}_{i^*\sim\rho}[\mathcal{L}(i^*, \boldsymbol{v})] = \sum_{i=1}^M \rho_i(1 - v_i) = 1 - \sum_{i=1}^M \rho_i v_i\,,$$

which is minimized by any distribution $\boldsymbol{v}$ with support on $\operatorname{argmax}_{1 \leqslant i \leqslant M} \rho_i$ (and in particular by the uniform distribution on this set). Taking $\rho = \pi(\cdot \,|\, \boldsymbol{\ell}_1, \ldots, \boldsymbol{\ell}_t)$ to be the posterior distribution, it follows from Equation (67) that a Bayes decision rule is given by the Follow-the-Leader (FTL) rule, which returns the uniform distribution over the experts i^* with smallest loss $L_{i^*,t}$.

Lower bounding the risk of FTL. Since FTL is a Bayes rule, we have

$$\begin{aligned}
\inf_A \mathcal{R}_t(\pi, A) &= \frac{1}{M} \sum_{i^*=1}^{M} \mathcal{R}_t(i^*, \mathrm{FTL}) \\
&= \mathcal{R}_t(1, \mathrm{FTL}) && (68)\\
&\geqslant \frac{1}{2} \mathbb{P}_1(\exists i \in \{2, \ldots, M\}, L_{i,t} \leqslant L_{1,t}) && (69)\\
&\geqslant \frac{1}{2} \mathbb{P}_1(L_{1,t} \geqslant t/2, \exists i \in \{2, \ldots, M\}, L_{i,t} \leqslant t/2) \\
&= \frac{1}{2} \mathbb{P}_1(L_{1,t} \geqslant t/2) \big(1 - (1 - \mathbb{P}_1(L_{2,t} \leqslant t/2))^{M-1}\big) && (70)
\end{aligned}$$

where (68) comes from the symmetry of FTL over the indices, (69) is due to the fact that if there exists $i \neq 1$ with $L_{i,t} \leqslant L_{1,t}$ then FTL gives a weight of at most $\frac{1}{2}$ to 1, and (70) follows from the independence of the losses. Now, using the distribution of the losses under $\mathbb{P}_1$, inequality (70) implies that

$$\inf_A \mathcal{R}_t(\pi, A) \geqslant \frac{1}{4} \big(1 - (1 - P_t^{(\frac{1}{2}+\Delta)}(S_t \leqslant t/2))^{M-1}\big) \tag{71}$$

where $P_t^{(\theta)}$ is the distribution $\mathcal{B}(\theta)^{\otimes t}$ on $\{0,1\}^t$, and $S_t : \{0,1\}^t \to \mathbf{R}$ is the random variable $S_t(x_1, \ldots, x_t) = x_1 + \cdots + x_t$.

Concluding the proof with a Pinsker bound. It follows from the inequalities (65), (66) and (71) that

$$\sup_{1 \leqslant i^* \leqslant M} \mathbb{E}_{i^*}[R_T] \geqslant \frac{\Delta}{4} \sum_{t=0}^{T-1} \big(1 - (1 - P_t^{(\frac{1}{2}+\Delta)}(S_t \leqslant t/2))^{M-1}\big)\,,$$

so that, since $1 - p \leqslant e^{-p}$ for $p \in \mathbf{R}$,

$$\sup_{1 \leqslant i^* \leqslant M} \mathbb{E}_{i^*}[R_T] \geqslant \frac{\Delta}{4} \sum_{t=0}^{T-1} \Big\{1 - e^{-(M-1) P_t^{(\frac{1}{2}+\Delta)}(S_t \leqslant t/2)}\Big\}\,. \tag{72}$$

In order to conclude the proof, it suffices to show that at least $\Omega\big(\frac{\log M}{\Delta^2}\big)$ terms inside the sum in (72) are of order $\Omega(1)$. This amounts to showing that $P_t^{(\frac{1}{2}+\Delta)}(S_t \leqslant t/2) \gtrsim \frac{1}{M}$ for $t \lesssim \frac{\log M}{\Delta^2}$, *i.e.* (solving for M) $P_t^{(\frac{1}{2}+\Delta)}(S_t \leqslant t/2) \geqslant C e^{-C' t \Delta^2}$ for some universal constants C, C' independent of Δ.

To do this, we recall the following Pinsker bound (a direct consequence of Lemma 2.6 in [25]): if P, Q are two probability distributions on some measurable space, then

$$P(A) + Q(A^c) \geqslant \frac{1}{2} \exp(-\mathrm{KL}(P, Q))$$

for any event A. Applying this to $P = P_t^{(\frac{1}{2}+\Delta)}$, $Q = P_t^{(\frac{1}{2}-\Delta)}$ on $\{0,1\}^t$ and $A = \{S_t \leqslant t/2\}$:

$$P_t^{(\frac{1}{2}+\Delta)}(S_t \leqslant t/2) + P_t^{(\frac{1}{2}-\Delta)}(S_t > t/2) \geqslant \frac{1}{2} \exp(-\mathrm{KL}(P_t^{(\frac{1}{2}+\Delta)}, P_t^{(\frac{1}{2}-\Delta)}))\,. \tag{73}$$

Since $\mathrm{KL}(P_t^{(\frac{1}{2}+\Delta)}, P_t^{(\frac{1}{2}-\Delta)}) = t\mathrm{kl}\big(\frac{1}{2}+\Delta, \frac{1}{2}-\Delta\big)$ where $\mathrm{kl}(p,q) := \mathrm{KL}(\mathcal{B}(p), \mathcal{B}(q)) = p\log\frac{p}{q} + (1-p)\log\frac{1-p}{1-q}$, and since

$$P_t^{(\frac{1}{2}-\Delta)}(S_t > t/2) = P_t^{(\frac{1}{2}-\Delta)}(t - S_t < t/2) = P_t^{(\frac{1}{2}+\Delta)}(S_t < t/2) \leqslant P_t^{(\frac{1}{2}+\Delta)}(S_t \leqslant t/2)\,,$$

inequality (73) yields:

$$P_t^{(\frac{1}{2}+\Delta)}(S_t \leqslant t/2) \geqslant \frac{1}{4}\exp\left(-t\mathrm{kl}\left(\frac{1}{2}+\Delta, \frac{1}{2}-\Delta\right)\right) \geqslant \frac{1}{4}e^{-18t\Delta^2}\,. \tag{74}$$

where the second inequality comes from the fact $p, q \in (\frac{1}{3}, \frac{2}{3})$, $\frac{\partial^2}{\partial q^2}\mathrm{kl}(p,q) = \frac{p}{q^2} + \frac{1-p}{(1-q)^2} \leqslant \frac{1}{(1/3)^2}(p + (1-p)) = 9$, so that $\mathrm{kl}(p,q) \leqslant \frac{9}{2}(p-q)^2$ by Taylor's inequality (recall that $\Delta \in (0, \frac{1}{6})$).

Finally, let $t_0 = \left\lfloor \frac{\log(2(M-1))}{18\Delta^2} \right\rfloor$, so that $\frac{M-1}{4}e^{-18t\Delta^2} \geqslant \frac{1}{8}$ for $0 \leqslant t \leqslant t_0$ and $t_0 + 1 \geqslant \frac{\log(2(M-1))}{18\Delta^2} \geqslant \frac{\log M}{18\Delta^2}$. It follows from (72) and (74) that:

$$\begin{aligned}\sup_{1\leqslant i^*\leqslant M} \mathbb{E}_{i^*}[R_T] &\geqslant \frac{\Delta}{4}\sum_{t=0}^{T-1}\left\{1 - e^{-(M-1)\frac{1}{4}\exp(-18t\Delta^2)}\right\}\\ &\geqslant \frac{\Delta}{4}\sum_{t=0}^{t_0\wedge(T-1)}\left(1 - e^{-1/8}\right)\,, \qquad (75)\end{aligned}$$

which provides the announced bound (7) by recalling that $T \geqslant \frac{\log M}{18\Delta^2}$ by assumption and lower bounding $\frac{1-e^{-1/8}}{4\times 18} \geqslant \frac{1}{620}$.

References

[1] P. Auer, N. Cesa-Bianchi, and C. Gentile. Adaptive and self-confident on-line learning algorithms. *Journal of Computer and System Sciences*, 64(1):48–75, 2002.

[2] K. Azuma. Weighted sums of certain dependent random variables. *Tohoku Mathematical Journal, Second Series*, 19(3):357–367, 1967.

[3] P. L. Bartlett and S. Mendelson. Empirical minimization. *Probability Theory and Related Fields*, 135(3):311–334, 2006.

[4] S. Boucheron, O. Bousquet, and G. Lugosi. Theory of classification: A survey of some recent advances. *ESAIM: probability and statistics*, 9:323–375, 2005.

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

[6] N. Cesa-Bianchi and G. Lugosi. *Prediction, Learning, and Games*. Cambridge University Press, Cambridge, New York, USA, 2006.

[7] N. Cesa-Bianchi, Y. Mansour, and G. Stoltz. Improved second-order bounds for prediction with expert advice. *Machine Learning*, 66:321–352, 2007.

[8] A. Chernov and F. Zhdanov. Prediction with expert advice under discounted loss. In *International Conference on Algorithmic Learning Theory (ALT)*, pages 255–269, 2010.

[9] S. de Rooij, T. van Erven, P. Grünwald, and W. M. Koolen. Follow the leader if you can, hedge if you must. *Journal of Machine Learning Research*, 15:1281–1316, 2014.

[10] 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.

[11] P. Gaillard, G. Stoltz, and T. van Erven. A second-order bound with excess losses. In *Proceedings of the 27th Annual Conference on Learning Theory (COLT)*, pages 176–196, 2014.

[12] W. Hoeffding. Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association*, 58(301):13–30, 1963.

[13] W. M. Koolen, P. Grünwald, and T. van Erven. Combining adversarial guarantees and stochastic fast rates in online learning. In *Advances in Neural Information Processing Systems 29*, pages 4457–4465. Curran Associates, Inc., 2016.

[14] W. M. Koolen and T. van Erven. Second-order quantile methods for experts and combinatorial games. In *Proceedings of the 28th Annual Conference on Learning Theory (COLT)*, pages 1155–75, 2015.

[15] W. M. Koolen, T. van Erven, and P. D. Grünwald. Learning the learning rate for prediction with expert advice. In *Advances in Neural Information Processing Systems 27*, pages 2294–2302. Curran Associates, Inc., 2014.

[16] N. Littlestone and M. K. Warmuth. The weighted majority algorithm. *Information and computation*, 108(2):212–261, 1994.

[17] H. Luo and R. E. Schapire. Achieving all with no parameters: AdaNormalHedge. In *Proceedings of the 28th Annual Conference on Learning Theory (COLT)*, pages 1286–1304, 2015.

[18] E. Mammen and A. B. Tsybakov. Smooth discrimination analysis. *The Annals of Statistics*, 27(6):1808–1829, 1999.

[19] P. Massart and É. Nédélec. Risk bounds for statistical learning. *The Annals of Statistics*, 34(5):2326–2366, 2006.

[20] E. Moulines and F. Bach. Non-asymptotic analysis of stochastic approximation algorithms for machine learning. In *Advances in Neural Information Processing Systems 24*, pages 451–459. Curran Associates, Inc., 2011.

[21] A. Sani, G. Neu, and A. Lazaric. Exploiting easy data in online optimization. In *Advances in Neural Information Processing Systems 27*, pages 810–818. Curran Associates, Inc., 2014.

[22] Y. Seldin and G. Lugosi. An improved parametrization and analysis of the EXP3++ algorithm for stochastic and adversarial bandits. In *Proceedings of the 2017 Conference on Learning Theory (COLT)*, pages 1743–1759, 2017.

[23] Y. Seldin and A. Slivkins. One practical algorithm for both stochastic and adversarial bandits. In *Proceedings of the 31st International Conference on Machine Learning (ICML)*, pages 1287–1295, 2014.

[24] A. B. Tsybakov. Optimal aggregation of classifiers in statistical learning. *Annals of Statistics*, 32(1):135–166, 2004.

[25] A. B. Tsybakov. *Introduction to nonparametric estimation.* Springer, 2009.

[26] T. van Erven, P. D. Grünwald, N. A. Mehta, M. D. Reid, and R. C. Williamson. Fast rates in statistical and online learning. *Journal of Machine Learning Research*, 16(1):1793–1861, 2015.

[27] V. Vovk. A game of prediction with expert advice. *Journal of Computer and System Sciences*, 56(2):153–173, 1998.

[28] O. Wintenberger. Optimal learning with Bernstein online aggregation. *Machine Learning*, 106(1):119–141, 2017.