On the optimality of the Hedge algorithm in the stochastic regime

Abstract. In this chapter, 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 remarkably 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. Moreover, our analysis exhibits qualitative differences with other versions of the Hedge algorithm, such as the fixed-horizon variant (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 determine the intrinsic limitations of anytime Hedge in the stochastic case, and discuss the improvements provided by more adaptive algorithms.

Contents

4.1   Introduction

The standard setting of *prediction with expert advice* (Littlestone and Warmuth, 1994; Freund and Schapire, 1997; Vovk, 1998; Cesa-Bianchi and Lugosi, 2006) aims to provide sound strategies for sequential prediction that combine the forecasts from different sources. More precisely, in the so-called *Hedge problem* (Freund and Schapire, 1997), at each round the learner has to output a probability distribution on a finite set of *experts* $\{1, \dots, M\}$; 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

173

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 (Cesa-Bianchi and Lugosi, 2006).

In this setting, arguably the simplest and most standard strategy is the *Hedge algorithm* (Freund and Schapire, 1997), also called the *exponentially weighted averaged forecaster* (Cesa-Bianchi and Lugosi, 2006). 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 $T \geqslant 1$.

While worst-case regret bounds are robust and always valid, they turn out to be overly pessimistic in some situations. A recent line of research (Cesa-Bianchi et al., 2007; de Rooij et al., 2014; Gaillard et al., 2014; Koolen et al., 2014; Sani et al., 2014; Koolen and van Erven, 2015; Luo and Schapire, 2015) designs algorithms that combine $O(\sqrt{T \log M})$ worst-case 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 "best" 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 (Cesa-Bianchi et al., 2007; de Rooij et al., 2014; Koolen et al., 2014; Gaillard et al., 2014), or on more sophisticated strategies (Koolen and van Erven, 2015; Luo and Schapire, 2015). As shown by Gaillard et al. (2014) (see also Koolen et al. 2016), 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 (Cesa-Bianchi et al., 2007), 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 (de Rooij et al., 2014; Gaillard et al., 2014; Koolen et al., 2014; Sani et al., 2014). However, to the best of our knowledge, existing work on the Hedge problem 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 (Littlestone and Warmuth, 1994; Freund and Schapire, 1997; Cesa-Bianchi and Lugosi, 2006), 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 work fills a gap in the existing literature by providing an analysis of the standard Hedge algorithm in the stochastic setting. We show that 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 dedicated tuning* for the easier instance, which might be surprising at first sight. 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 exhibits a strong difference between the performances of the anytime and the fixed-horizon variants of the Hedge algorithm.

Given the aforementioned adaptivity of Decreasing Hedge, one may wonder whether there is in fact any benefit in using more sophisticated algorithms in the stochastic regime. We answer this question affirmatively, by considering a more refined measure of complexity of a stochastic instance than the gap Δ. Specifically, we show that Decreasing Hedge does not admit improved regret under Bernstein conditions, which are standard low-noise conditions from statistical learning (Mammen and Tsybakov, 1999; Tsybakov, 2004; Bartlett and Mendelson, 2006). By contrast, it was shown by Koolen et al. (2016) that algorithms which satisfy some adaptive adversarial regret bound achieve improved regret under Bernstein conditions. Finally, we characterize the behavior of Decreasing Hedge in the stochastic regime, by showing that its eventual regret on *any* stochastic instance is governed by the gap Δ.

Related work. In the bandit setting, where the feedback only consists of the loss of the selected action, there has also been some interest in "best-of-both-worlds" algorithms that combine optimal $O(\sqrt{MT})$ worst-case regret in the adversarial regime with improved $O(M \log T)$ regret (up to logarithmic factors) in the stochastic case (Bubeck and Slivkins, 2012; Seldin and Slivkins, 2014; Auer and Chiang, 2016). In particular, Seldin and Slivkins (2014); Seldin and Lugosi (2017) showed that by augmenting the standard EXP3 algorithm for the adversarial regime (an analogue of Hedge with $\Theta(1/\sqrt{t})$ learning rate) with a special-purpose gap detection mechanism, one can achieve poly-logarithmic regret in the stochastic case. This result is strengthened in some recent follow-up work (Zimmert and Seldin, 2019; Zimmert et al., 2019), which appeared since the completion of the first version of the present work, that obtains optimal regret in the stochastic and adversarial regimes through a variant of the Follow-The-Regularized-Leader (FTRL) algorithm with $\Theta(1/\sqrt{t})$ learning rate and a proper regularizer choice. This result can be seen as an analogue in the bandit case of our upper bound for Decreasing Hedge. On the other hand, in the bandit setting, the hardness of an instance is essentially characterized by the gap Δ (Bubeck and Cesa-Bianchi, 2012); in particular, the Bernstein condition, which depends on the correlations between the losses of the experts, cannot be exploited under bandit feedback, where one only observes one arm at each round. Hence, it appears that the negative part of our results (on the limitations of Hedge) does not have an analogue in the bandit case.

A similar adaptivity result for FTRL with decreasing $\Theta(1/\sqrt{t})$ learning rate has been observed in a different context by Huang et al. (2017). Specifically, it is shown that, in the case of online linear optimization on a Euclidean ball, FTRL with squared norm regularizer and learning rate $\Theta(1/\sqrt{t})$ achieves $O(\log T)$ regret when the loss vectors are i.i.d. This result is an analogue of our upper bound for Hedge, since this algorithm corresponds to FTRL on the simplex with entropic regularizer (Cesa-Bianchi and Lugosi, 2006; Hazan, 2016). On the other hand, the simplex lacks the curvature of the Euclidean ball, which is important to achieve small regret; here, the improved regret is ensured by a condition on the distribution, namely the existence of a gap Δ. Our lower bound for Hedge shows that this condition is necessary, thereby characterizing the long-term regret of FTRL on the simplex with entropic regularizer. In the case of the Euclidean ball with squared norm regularizer, the norm of the expected loss vector appears to play a similar role, as shown by the upper bound from Huang et al. (2017).

Outline. We define the setting of prediction with expert advice and the Hedge algorithm in Section 4.2, and we recall herein its standard worst-case regret bound. In Section 4.3, 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 4.1 an upper bound on the regret of order $(\log M)/\Delta$ for Decreasing Hedge. In Proposition 4.2, we prove that the rate $(\log M)/\Delta$ cannot be improved in this setting. In Theorem 4.2 and Corollary 4.1, 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 4.3, 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 4.4, we discuss the advantages of adaptive Hedge algorithms, and explain what the limitations of Decreasing Hedge are compared to such versions. We include numerical illustrations of our theoretical findings in Section 4.5, conclude in Section 4.6 and provide the proofs in Section 4.7.

4.2 The expert problem and the Hedge algorithm

In the Hedge setting, also called *decision-theoretic online learning* (Freund and Schapire, 1997), 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 $\widehat{\ell}_t = \boldsymbol{v}_t^\top \boldsymbol{\ell}_t$.

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

$$R_T = \sum_{t=1}^{T} \widehat{\ell}_t - \min_{1 \leqslant i \leqslant M} \sum_{t=1}^{T} \ell_{i,t} \tag{4.1}$$

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

$$v_{i,t} = \frac{e^{-\eta_t L_{i,t-1}}}{\sum_{j=1}^{M} e^{-\eta_t L_{j,t-1}}} \tag{4.2}$$

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 $\widehat{L}_T := \sum_{t=1}^{T} \widehat{\ell}_t$ and $R_{i,T} = \widehat{L}_T - L_{i,T}$ the regret with respect to expert i. We consider in this chapter the following variants of Hedge, where $c_0 > 0$ is a constant.

Decreasing Hedge (Auer et al., 2002). This is Hedge with the sequence of learning rates $\eta_t = c_0 \sqrt{\log(M)/t}$.

Constant Hedge (Littlestone and Warmuth, 1994). Given a finite time horizon $T \geqslant 1$, this is Hedge with constant learning rate $\eta_t = c_0 \sqrt{\log(M)/T}$.

Hedge with doubling trick (Cesa-Bianchi et al., 1997; Cesa-Bianchi and Lugosi, 2006). 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

$$v_{i,t} = \frac{\exp(-\eta_t \sum_{s=T_k}^{t-1} \ell_{i,s})}{\sum_{j=1}^{M} \exp(-\eta_t \sum_{s=T_k}^{t-1} \ell_{j,s})}, \tag{4.3}$$

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

Let us recall the following standard regret bound for the Hedge algorithm from Chernov and Zhdanov (2010).

Proposition 4.1. *Let $\eta_1, \eta_2, \dots$ be a decreasing sequence of learning rates. The Hedge algorithm (4.2) satisfies the following regret bound:*

$$R_T \leqslant \frac{1}{\eta_T} \log M + \frac{1}{8} \sum_{t=1}^{T} \eta_t \,. \tag{4.4}$$

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 (4.4) holds for every sequence of losses $\ell_1, \ell_2, \dots$, 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 (Cesa-Bianchi and Lugosi, 2006). 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, Cesa-Bianchi and Lugosi, 2006) may achieve constant regret in easier instances, even though they lack regret guarantees in the adversarial regime. We show in Section 4.3 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.

4.3 Regret of 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.

4.3.1 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 $\boldsymbol{\ell}_1, \boldsymbol{\ell}_2, \ldots$ 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 Gaillard et al. (2014); Luo and Schapire (2015), 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 $\Delta = \min_{i \neq i^*} \mathbb{E}[\ell_{i,t} - \ell_{i^*,t}]$, where $i^* = \arg\min_i \mathbb{E}[\ell_{i,t}]$.

We show below that, despite the fact that Decreasing Hedge is designed for the worst-case setting described in Section 4.2, it is able to adapt to the easier problem considered here, Indeed, Theorem 4.1 shows that Decreasing Hedge achieves a *constant*, and in fact *optimal* (by Proposition 4.2 below) regret bound in this setting, in spite of its "conservative" learning rate.

With the exception of the high-probability bound of Corollary 4.1, the upper and lower bounds in the stochastic case are stated for the *pseudo-regret* $\mathcal{R}_T = \mathbb{E}[R_{i^*,T}]$ (similar bounds hold for the the expected regret $\mathbb{E}[R_T]$, since $\mathcal{R}_T \leqslant \mathbb{E}[R_T]$ and by Remark 4.3 in Section 4.7.1).

Theorem 4.1. *Let $M \geqslant 3$. Assume that the loss vectors $\boldsymbol{\ell}_1, \boldsymbol{\ell}_2, \ldots$ are i.i.d. random variables, where $\boldsymbol{\ell}_t = (\ell_{i,t})_{1 \leqslant i \leqslant M}$. Also, assume that there exists $i^* \in \{1, \ldots, M\}$ and $\Delta > 0$ such that*

$$\mathbb{E}[\ell_{i,t} - \ell_{i^*,t}] \geqslant \Delta \tag{4.5}$$

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

$$\mathcal{R}_T \leqslant \frac{4 \log M + 25}{\Delta} . \tag{4.6}$$

The proof of Theorem 4.1 is given in Section 4.7.1. Theorem 4.1 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 (Gaillard et al., 2014; Luo and Schapire, 2015). 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 (Cesa-Bianchi et al., 2007) or AdaHedge (de Rooij et al., 2014), 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 4.1 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 4.1 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 4.1, 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. with high probability, 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, the learning rate $\eta_t \asymp \sqrt{(\log M)/t}$ 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 $\sum_{t \geqslant 1} e^{-c\sqrt{t}} = O(\frac{1}{c^2})$ for $c > 0$.

Remark 4.1. The fact that $\sum_{t \geqslant 1} e^{-c\sqrt{t}} = O(1/c^2)$ is also used in the analysis of the EXP3++ bandit algorithm (Seldin and Slivkins, 2014, Lemma 10). In the expert setting considered here, summing the contribution all experts (which suffices in the bandit setting to obtain the correct order of regret) would yield a significantly suboptimal $O(M/\Delta)$ regret bound, with a linear dependence on the number of experts 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 obtain the optimal rate $(\log M)/\Delta$.

We complement Theorem 4.1 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.

Proposition 4.2. *Let $\Delta \in (0, \frac{1}{4})$, $M \geqslant 4$ and $T \geqslant (\log M)/(16\Delta^2)$. For any algorithm for the Hedge setting, there exists an i.i.d. distribution over sequences of losses $(\boldsymbol{\ell}_t)_{t \geqslant 1}$ such that:*

- *there exists $i^* \in \{1, \ldots, M\}$ such that, for any $i \neq i^*$, $\mathbb{E}[\ell_{i,t} - \ell_{i^*,t}] \geqslant \Delta$;*

- *the pseudo-regret of the algorithm satisfies:*

$$\mathcal{R}_T \geqslant \frac{\log M}{256\Delta}. \tag{4.7}$$

The proof of Proposition 4.2 is given in Section 4.7.2. Proposition 4.2 generalizes the well-known minimax lower bound of $\Theta(\sqrt{T \log M})$, which is recovered by taking $\Delta \asymp \sqrt{(\log M)/T}$.

4.3.2 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 4.1), 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 by Seldin and Slivkins (2014) in the bandit case, where the leading expert linearly outperforms the others after some time. As Theorem 4.2 shows, essentially the same regret guarantee can be obtained in this case, up to an additional $\log(\Delta^{-1})/\Delta$ term. Theorem 4.2 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 (Auer et al., 2002; Cesa-Bianchi et al., 2007; de Rooij et al., 2014). In what follows, we will assume $M \geqslant 3$ for convenience; similar results holds for $M = 2$.

Theorem 4.2. *Let $M \geqslant 3$. Assume that there exists $\tau_0 \geqslant 1$, $\Delta \in (0,1)$ and $i^* \in \{1, \ldots, M\}$ such that, for every $t \geqslant \tau_0$ and $i \neq i^*$, one has*

$$L_{i,t} - L_{i^*,t} \geqslant \Delta t. \tag{4.8}$$

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 $c_0 > 0$;*

- *it 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

$$R_T \leqslant c_1 \sqrt{\tau_0 \log M} + \frac{c_2 \log M + c_3 \log \Delta^{-1} + c_4}{\Delta} \tag{4.9}$$

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 4.2 is the same as that of Theorem 4.1, 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 4.2, we can recover the guarantee of Theorem 4.1 (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 4.2 and Corollary 4.1 are provided in Section 4.7.3.

Corollary 4.1. *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*

$$\mathbb{E}\left[\ell_{i,t} - \ell_{i^*,t} | \mathcal{F}_{t-1}\right] \geqslant \Delta \tag{4.10}$$

for every $i \neq i^*$ and every $t \geqslant 1$. Then, for any Hedge algorithm satisfying the conditions of Theorem 4.2, and every $T \geqslant 1$:

$$\mathcal{R}_T \leqslant (5c_1 + 2c_2)\frac{\log M}{\Delta} + 2c_3\frac{\log \Delta^{-1}}{\Delta} + \frac{2c_4}{\Delta}, \tag{4.11}$$

with c_1, c_2, c_3, c_4 as in Theorem 4.2. In addition, for every $\varepsilon \in (0,1)$, we have

$$R_T \leqslant \left(c_1\sqrt{8} + 2c_2\right)\frac{\log M}{\Delta} + c_1\frac{\sqrt{8\log M \log \varepsilon^{-1}}}{\Delta} + 2c_3\frac{\log \Delta^{-1}}{\Delta} + \frac{2c_4}{\Delta} \tag{4.12}$$

with probability at least $1 - \varepsilon$.

4.3.3 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 4.3.1 and 4.3.2, is not shared by the closely related Constant Hedge and Hedge with doubling trick, despite the fact that they both achieve the minimax optimal worst-case $O(\sqrt{T \log M})$ regret. Proposition 4.3 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 4.3. *Let $T \geqslant 1$, $M \geqslant 2$, and consider the experts $i = 1, \ldots, M$ with losses $\ell_{1,t} = 0$, $\ell_{i,t} = 1$ ($1 \leqslant t \leqslant T$, $2 \leqslant i \leqslant M$). Then, the pseudo-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:*

$$\mathcal{R}_T \geqslant \min\Big(\frac{\sqrt{T \log M}}{3 c_0}, \frac{T}{3}\Big). \tag{4.13}$$

In addition, Hedge with doubling trick (4.3) also suffers a pseudo-regret satisfying

$$\mathcal{R}_T \geqslant \min\Big(\frac{\sqrt{T \log M}}{6 c_0}, \frac{T}{12}\Big). \tag{4.14}$$

The proof of Proposition 4.3 is given in Section 4.7.4. Although Hedge with a doubling trick is recognized to be overly conservative and only suitable for worst-case scenarios Cesa-Bianchi and Lugosi, 2006 (especially due to its periodic restarts, after which it discards past observations), to the best of our knowledge Proposition 4.3 (together with Theorem 4.1) 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.

4.4 Limitations of Decreasing Hedge in the stochastic case

In this section, we explore the limitations of the simple Decreasing Hedge algorithm in the stochastic regime, and exhibit situations where it performs worse than more sophisticated algorithms. The starting observation is that the sub-optimality gap Δ is a rather brittle measure of "hardness" of a stochastic instance, which does not fully reflect the achievable rates. We therefore consider the following fast-rate condition from statistical learning, which refines the sub-optimality gap as a measure of complexity of a stochastic instance.

Definition 4.1 (Bernstein condition). Assume that the losses $\boldsymbol{\ell}_1, \boldsymbol{\ell}_2, \ldots$ are the realization of a stochastic process. Denote $\mathcal{F}_t = \sigma(\boldsymbol{\ell}_1, \ldots, \boldsymbol{\ell}_t)$ the σ-algebra generated by $\boldsymbol{\ell}_1, \ldots, \boldsymbol{\ell}_t$. 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 $i \neq i^*$,

$$\mathbb{E}[(\ell_{i,t} - \ell_{i^*,t})^2 | \mathcal{F}_{t-1}] \leqslant B \mathbb{E}[\ell_{i,t} - \ell_{i^*,t} | \mathcal{F}_{t-1}]^\beta. \tag{4.15}$$

The Bernstein condition (Bartlett and Mendelson, 2006), a generalization of the Tsybakov margin condition (Tsybakov, 2004; Mammen and Tsybakov, 1999), 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 van Erven et al. (2015) for a discussion of fast rates conditions. The Bernstein condition (4.15) quantifies the "easiness" of a stochastic instance, and generalizes the gap condition considered in the previous section (see Example 4.1 below). Roughly speaking, it states that good experts (with near-optimal expected loss) are highly correlated with the best expert. In the examples below, we assume that the loss vectors $\boldsymbol{\ell}_1, \boldsymbol{\ell}_2, \ldots$ are i.i.d.

Example 4.1 (Gap implies Bernstein). If $\Delta_i = \mathbb{E}[\ell_{i,t} - \ell_{i^*,t}] \geqslant \Delta$ for $i \neq i^*$, then the $(1, \frac{1}{\Delta})$-Bernstein condition holds (Koolen et al., 2016, Lemma 4). Furthermore, letting $\alpha = \mathbb{E}[\ell_{i^*,t}]$

denote the expected loss of the best expert, the $(1, 1 + \frac{2\alpha}{\Delta})$-Bernstein condition holds. 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 $u, v \in [0, 1]$):

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

which establishes the claim since $\Delta_i \geqslant \Delta$. This provides an improvement when α is small.

Example 4.2 (Bernstein without a gap)**.** Let P be a distribution on $\mathcal{X} \times \{0, 1\}$, where $\mathcal{X}$ is some measurable space. Assume that $(X_1, Y_1), (X_2, Y_2) \ldots$ are i.i.d. samples from P, and that the experts $i \in \{1, \ldots, M\}$ correspond to classifiers $f_i : \mathcal{X} \to \{0, 1\}$: $\ell_{i,t} = \mathbf{1}(f_i(X_t) \neq Y_t)$, and that expert i^* is the Bayes classifier: $f_{i^*}(X) = \mathbf{1}(\eta(X) \geqslant 1/2)$, where $\eta(X) = \mathbb{P}(Y = 1 | X)$. Tsybakov's low noise condition (Tsybakov, 2004), namely $\mathbb{P}(|2\eta(X) - 1| \leqslant t) \leqslant C t^\kappa$ 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., Boucheron et al., 2005). In addition, under the Massart condition (Massart and Nédélec, 2006) that $|\eta(X) - 1/2| \geqslant c > 0$, the $(1, 1/(2c))$-Bernstein condition holds. Note that these conditions may hold even with an arbitrarily small sub-optimality gap Δ, since the f_i, $i \neq i^*$, may be arbitrary.

Theorem 4.3 below shows that Decreasing Hedge fails to achieve improved rates under Bernstein conditions.

Theorem 4.3. *For every $T \geqslant 1$, there exists a $(1, 1)$-Bernstein stochastic instance on which the pseudo-regret of the Decreasing Hedge algorithm with $\eta_t = c_0 \sqrt{(\log M)/t}$ satisfies*

$$\mathcal{R}_T \geqslant \frac{1}{3} \min \left(\frac{1}{c_0} \sqrt{T \log M}, T\right).$$

The proof of Theorem 4.3 is given in Section 4.7.6. By contrast, it was shown by Koolen et al. (2016) (and implicitly used by Gaillard et al., 2014) that some adaptive algorithms with data-dependent regret bounds enjoy improved regret under the Bernstein condition. For the sake of completeness, we state this fact in Proposition 4.4 below, which corresponds to Koolen et al. (2016, Theorem 2), but where the dependence on B is made explicit. We also only provide a bound in expectation, which considerably simplifies the proof. The proof of Proposition 4.4, which uses the same ideas as Gaillard et al. (2014, Theorem 11), is provided in Section 4.7.5.

Proposition 4.4. *Consider an algorithm for the Hedge problem which satisfies the following regret bound: for every $i \in \{1, \ldots, M\}$,*

$$R_{i,T} \leqslant C_1 \sqrt{(\log M) \sum_{t=1}^{T} (\widehat{\ell}_t - \ell_{i,t})^2} + C_2 \log M \tag{4.16}$$

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

$$\mathcal{R}_T \leqslant C_3 (B \log M)^{\frac{1}{2-\beta}} T^{\frac{1-\beta}{2-\beta}} + C_4 \log M \tag{4.17}$$

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

The data-dependent regret bound (4.16), a "second-order" bound, is satisfied by adaptive algorithms such as Adapt-ML-Prod (Gaillard et al., 2014) and Squint (Koolen and van Erven, 2015). A slightly different variant of second-order regret bounds, which depends on some cumulative variance of the losses across experts, has been considered by Cesa-Bianchi et al. (2007); de Rooij et al. (2014), and is achieved by Hedge algorithms with a data-dependent tuning of the learning rate. Second-order bounds refine so-called *first-order* bounds (Cesa-Bianchi et al., 1997; Auer et al., 2002; Cesa-Bianchi and Lugosi, 2006), which are adversarial regret bounds that scale as $O(\sqrt{L_T^* \log M} + \log M)$, where L_T^* denotes the cumulative loss of the best expert. While first-order bounds may still scale as the worst-case $O(\sqrt{T \log M})$ rate in a typical stochastic instance (where the best expert has a positive expected loss), second-order algorithms are known to achieve constant $O((\log M)/\Delta)$ regret in the stochastic case with gap Δ (Gaillard et al., 2014; Koolen and van Erven, 2015).

Theorem 4.3, in light of Proposition 4.4, clarifies where the advantage of second-order algorithms compared to Decreasing Hedge lies: unlike the latter, they can exploit Bernstein conditions on the losses. The contrast is most apparent for Bernstein instances with $\beta = 1$. By Example 4.1, the existence of a gap Δ implies that the $(1, B)$-Bernstein condition holds with $B \leqslant \frac{1}{\Delta}$. However, as shown by Example 4.2, B can in fact be much smaller than Δ, in which case the regret bound (4.17) satisfied by second-order algorithms, namely $O(B \log M)$, significantly improves over the upper bound of $O((\log M)/\Delta)$ of Decreasing Hedge from Theorem 4.1. Theorem 4.3 provides an instance where the difference does occur, in the most pronounced case where $B = 1$, so that second-order algorithms enjoy small $O(\log M)$ regret, while Decreasing Hedge suffers $\Theta(\sqrt{T \log M})$ regret.

Remark 4.2. The advantage of larger learning rates on some stochastic instances may be understood intuitively as follows. Consider an instance with B small but small gap Δ. 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, near-optimal experts (with small gap Δ_i) are ruled out late (after $(\log M)/\Delta_i^2$ rounds). On the other hand, the Bernstein assumption entails that those experts are highly correlated with the best expert, hence 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.

Setting the Bernstein condition aside, we conclude by investigating the intrinsic limitations of Decreasing Hedge in the stochastic setting. Indeed, it is natural to ask whether Decreasing Hedge can exploit some other regularity of a stochastic instance, apart from the gap Δ. Theorem 4.4 shows that this is in fact not the case.

Theorem 4.4. *For every i.i.d. (over time) stochastic instance with a unique best expert*

$$i^* = \operatorname*{arg\,min}_{1 \leqslant i \leqslant M} \mathbb{E}[\ell_{i,t}],$$

the pseudo-regret of Decreasing Hedge (with $c_0 \geqslant 1$) satisfies

$$\mathcal{R}_T \geqslant \frac{1}{450 c_0^4 (\log M)^2 \Delta}$$

for $T \geqslant \frac{1}{4\Delta^2}$, where $\Delta := \inf_{i \neq i^*} \mathbb{E}[\ell_{i,t} - \ell_{i^*,t}]$.

Theorem 4.4 shows (together with the upper bound of Theorem 4.1) 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.

4.5 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 $c_0 = \sqrt{8}$, `adahedge` is the AdaHedge algorithm from de Rooij et al. (2014), which is a variant of the Hedge algorithm with a data-dependent tuning of the learning rate η_t (based on $\boldsymbol{\ell}_1, \ldots, \boldsymbol{\ell}_{t-1}$). As shown in the note Koolen (2018), AdaHedge also benefits from Bernstein conditions. A related algorithm, namely Hedge with second-order tuning of the learning rate (Cesa-Bianchi et al., 2007), performed similarly to AdaHedge on the examples considered below, and was therefore not included. `FTL` is Follow-the-Leader (Cesa-Bianchi and Lugosi, 2006) 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 4.1 the cumulative regrets of the considered algorithms in four examples. The results for the stochastic instances (a), (b) and (c) described below are averaged over 50 trials.

(*a*) *Stochastic instance with a gap.* This is the standard instance considered here. The losses are drawn independently from Bernoulli distributions (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 4.1a 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 4.1 (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, Ada-Hedge and FTL achieve better regret than the two conservative Hedge variants (Figure 4.1b). 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
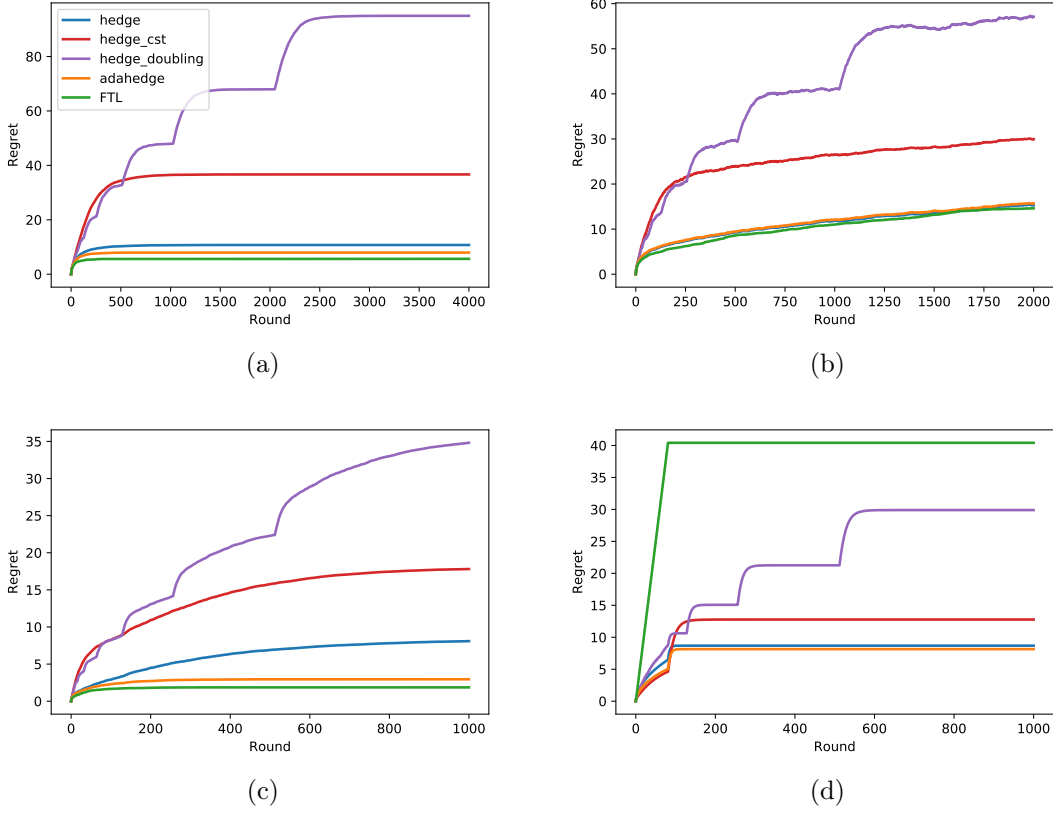
Figure 4.1: Cumulative regret of Hedge algorithms on four examples, see text for a precise description and discussion about the results. (a) Stochastic instance with a gap; (b) "Hard" stochastic instance; (c) Small loss for the best expert; (d) Adversarial with a gap instance.

improved regret bounds when the leading expert has small loss. 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 4.2. 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.

4.6 Conclusion

In this chapter, 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 that, 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 one 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 link with stochastic optimization. Our results have a similar flavor to a well-known result (Moulines and Bach, 2011; Bach, 2014) in 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 iterate averaging, and it does not come from a regret bound. 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.

4.7 Proofs

We now provide the proofs of the results from the previous sections, by order of appearance in the text.

4.7.1 Proof of Theorem 4.1

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 4.1) shows that for $1 \leqslant T \leqslant t_0$:

$$R_{i^*,T} \leqslant \sqrt{T\log M} \leqslant \sqrt{t_0\log M} \leqslant \sqrt{\log M} + \frac{2\sqrt{2}\log M}{\Delta} \leqslant \frac{4\log M}{\Delta} \qquad (4.18)$$

(since $\log M \geqslant 1$ as $M \geqslant 3$, $\Delta \leqslant 1$ and $2\sqrt{2} \leqslant 3$), which establishes (4.6) for $T \leqslant t_0$. In order to prove (4.6) for $T \geqslant t_0 + 1$, we start by decomposing the regret with respect to i^* as

$$R_{i^*,T} = \widehat{L}_T - L_{i^*,T} = \widehat{L}_{t_0} - L_{i^*,t_0} + \sum_{t=t_0+1}^{T} (\widehat{\ell}_t - \ell_{i^*,t}). \qquad (4.19)$$

Since $\widehat{L}_{t_0} - L_{i^*,t_0} \leqslant R_{t_0}$ is controlled by (4.18), it remains to upper bound the second term in (4.19). First, for every $t \geqslant t_0 + 1$,

$$\widehat{\ell}_t - \ell_{i^*,t} = \sum_{i \neq i^*} v_{i,t}(\ell_{i,t} - \ell_{i^*,t}). \tag{4.20}$$

Since $\boldsymbol{\ell}_t$ is independent of $\boldsymbol{v}_t$ (which is $\sigma(\boldsymbol{\ell}_1, \dots, \boldsymbol{\ell}_{t-1})$-measurable), taking the expectation in (4.20) yields, denoting $\Delta_i = \mathbb{E}[\ell_{i,t} - \ell_{i^*,t}]$,

$$\mathbb{E}[\widehat{\ell}_t - \ell_{i^*,t}] = \sum_{i \neq i^*} \Delta_i \mathbb{E}[v_{i,t}]. \tag{4.21}$$

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

$$\mathbb{P}\left(L_{i,t-1} - L_{i^*,t-1} < \frac{\Delta_i(t-1)}{2}\right) = \mathbb{P}\left(\sum_{s=1}^{t-1} Z_{i,s} > \frac{\Delta_i(t-1)}{2}\right)$$
$$\leqslant e^{-\frac{t-1}{2}(\Delta_i/2)^2}$$
$$= e^{-(t-1)\Delta_i^2/8}. \tag{4.22}$$

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

$$v_{i,t} = \frac{e^{-\eta_t(L_{i,t-1} - L_{i^*,t-1})}}{1 + \sum_{j \neq i^*} e^{-\eta_t(L_{j,t-1} - L_{i^*,t-1})}}$$
$$\leqslant e^{-2\sqrt{(\log M)/t} \times \Delta_i(t-1)/2}$$
$$\leqslant e^{-\Delta_i\sqrt{(t-1)(\log M)/2}} \tag{4.23}$$

since $t \leqslant 2(t-1)$. It follows from (4.23) and (4.22) that, for $t \geqslant t_0 + 1 \geqslant 2$,

$$\mathbb{E}[v_{i,t}] \leqslant \mathbb{P}\left(L_{i,t-1} - L_{i^*,t-1} > \frac{\Delta_i(t-1)}{2}\right) + e^{-\Delta_i\sqrt{(t-1)(\log M)/2}}$$
$$\leqslant e^{-(t-1)\Delta_i^2/8} + e^{-\Delta_i\sqrt{(t-1)(\log M)/2}}. \tag{4.24}$$

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

$$\Delta_i e^{-(t-1)\Delta_i^2/8} = \frac{2}{\sqrt{t-1}} f_2\left(\frac{\sqrt{t-1}\Delta_i}{2}\right) \leqslant \frac{2}{\sqrt{t-1}} f_2\left(\frac{\sqrt{t-1}\Delta}{2}\right) = \Delta e^{-(t-1)\Delta^2/8} \tag{4.25}$$

provided that $\frac{\sqrt{t-1}\Delta}{2} \geqslant 1$, i.e. $t \geqslant 1 + \frac{4}{\Delta^2}$, which is the case since $t \geqslant t_0 + 1 \geqslant 1 + \frac{8\log M}{\Delta^2}$. Likewise,

$$\Delta_i e^{-\Delta_i\sqrt{(t-1)(\log M)/2}} \leqslant \Delta e^{-\Delta\sqrt{(t-1)(\log M)/2}} \tag{4.26}$$

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 (4.21), (4.24), (4.25) and (4.26) that for every $t \geqslant t_0 + 1$:

$$\mathbb{E}[\widehat{\ell}_t - \ell_{i^*,t}] \leqslant M\Delta e^{-(t-1)\Delta^2/8} + M\Delta e^{-\Delta\sqrt{(t-1)(\log M)/2}}$$
$$= \left(Me^{-t_0\Delta^2/8}\right)\left(\Delta e^{-(t-t_0-1)\Delta^2/8}\right) + \left(Me^{-\Delta\sqrt{(t-1)(\log M)/8}}\right)\left(\Delta e^{-\Delta\sqrt{(t-1)(\log M)/8}}\right)$$
$$\leqslant \Delta e^{-(t-t_0-1)\Delta^2/8} + \Delta e^{-\Delta\sqrt{(t-1)/8}} \tag{4.27}$$

where inequality (4.27) 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 (4.27) yields, for every $T \geqslant t_0 + 1$,

$$
\begin{aligned}
\mathbb{E}[\sum_{t=t_0+1}^{T}(\ell_t - \ell_{i^*,t})] &\leqslant \sum_{t=t_0+1}^{T}\left\{\Delta e^{-(t-t_0-1)\Delta^2/8} + \Delta e^{-\Delta\sqrt{(t-1)/8}}\right\} \\
&\leqslant \Delta\sum_{t\geqslant 0}e^{-t\Delta^2/8} + \Delta\sum_{t\geqslant 1}e^{-(\Delta/\sqrt{8})\sqrt{t}} \\
&\leqslant \Delta\left(1 + \frac{8}{\Delta^2}\right) + \Delta\times\frac{2}{(\Delta/\sqrt{8})^2} \qquad (4.28) \\
&\leqslant \frac{25}{\Delta} \qquad\qquad\qquad\qquad\qquad\qquad (4.29)
\end{aligned}
$$

where inequality (4.28) comes from Lemma 4.1 below. Finally, combining inequalities (4.18) and (4.28) yields the pseudo-regret bound $\mathcal{R}_T \leqslant \frac{4\log M+25}{\Delta}$.

Lemma 4.1. *For every $\alpha > 0$,*

$$
\sum_{t\geqslant 1}e^{-\alpha t} \leqslant \frac{1}{\alpha} \qquad\qquad\qquad\qquad (4.30)
$$

$$
\sum_{t\geqslant 1}e^{-\alpha\sqrt{t}} \leqslant \frac{2}{\alpha^2}. \qquad\qquad\qquad (4.31)
$$

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

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

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

Remark 4.3. While the upper bound of Theorem 4.1 is stated for the pseudo-regret $\mathcal{R}_T$, a similar upper bound holds for the expected regret $\mathbb{E}[R_T]$. Indeed, under the assumptions of Theorem 4.1, for every $T \geqslant \frac{4\log M}{\Delta^2}$, we have $\mathbb{E}[R_T] \leqslant \mathcal{R}_T + \frac{1.1}{\Delta}$.

Proof. Note that $\mathbb{E}[R_T] - \mathcal{R}_T = \mathbb{E}[L_{i^*,T} - \min_{1\leqslant i\leqslant T}L_{i,T}]$. 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

$$
\begin{aligned}
\mathbb{P}\left(L_{i^*,T} - \min_{1\leqslant i\leqslant T}L_{i,T} \geqslant a\right) &\leqslant \sum_{i\neq i^*}\mathbb{P}\left(L_{i^*,T} - L_{i,T} + \Delta_i T \geqslant \Delta_i T + a\right) \\
&\leqslant \sum_{i\neq i^*}e^{-(\Delta_i T+a)^2/(2T)} \qquad\qquad\qquad\qquad (4.32) \\
&\leqslant Me^{-T\Delta^2/2}e^{-a^2/(2T)} \\
&\leqslant e^{-T\Delta^2/4}e^{-a^2/(2T)}, \qquad\qquad\qquad\qquad (4.33)
\end{aligned}
$$

where inequality (4.33) 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

$$
\begin{aligned}
\mathbb{E}\left[L_{i^*,T} - \min_{1\leqslant i \leqslant T} L_{i,T}\right] &= \int_0^\infty \mathbb{P}\left(L_{i^*,T} - \min_{1\leqslant i \leqslant T} L_{i,T} \geqslant a\right) \mathrm{d}a \\
&\leqslant e^{-T\Delta^2/4} \int_0^\infty e^{-a^2/(2T)} \mathrm{d}a \\
&= \sqrt{\frac{\pi}{2}} \cdot \sqrt{T} e^{-T\Delta^2/4} \\
&= \frac{\sqrt{\pi}}{\Delta}\left[\Delta\sqrt{T/2} \cdot e^{-(\Delta\sqrt{T/2})^2/2}\right] \\
&\leqslant \frac{\sqrt{\pi/e}}{\Delta}
\end{aligned}
\tag{4.34}
$$

where inequality (4.34) 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 $\sqrt{\pi/e} \leqslant 1.1$. $\qquad\square$

4.7.2 Proof of Proposition 4.2

Fix M, Δ and T as in Proposition 4.2. For $i^* \in \{1, \ldots, M\}$, 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 $\frac{1}{2} - \Delta$ if $i = i^*$ and $\frac{1}{2}$ otherwise; also, denote by $\mathbb{E}_{i^*}$ the expectation with respect to $\mathbb{P}_{i^*}$. 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 $(\boldsymbol{\ell}_1, \ldots, \boldsymbol{\ell}_{t-1})$ to an element of the probability simplex $\mathcal{P}_M \subset \mathbf{R}^M$ on $\{1, \ldots, M\}$. For any $i^* \in \{1, \ldots, M\}$, let $\mathcal{R}_T(i^*, \mathcal{A})$ denote the pseudo-regret of algorithm $\mathcal{A}$ under the distribution $\mathbb{P}_{i^*}$. Since $\boldsymbol{\ell}_t$ is independent of $\boldsymbol{v}_t$ under $\mathbb{P}_{i^*}$, we have

$$
\mathcal{R}_T(i^*, \mathcal{A}) = \sum_{t=1}^T \sum_{i \neq i^*} \mathbb{E}_{i^*}\left[v_{i,t}(\ell_{i,t} - \ell_{i^*,t})\right] = \Delta \sum_{t=1}^T \sum_{i \neq i^*} \mathbb{E}_{i^*}[v_{i,t}] = \Delta \sum_{t=1}^T \mathbb{E}_{i^*}[1 - v_{i^*,t}] \tag{4.35}
$$

with $\boldsymbol{v}_t := A_t(\boldsymbol{\ell}_1, \ldots, \boldsymbol{\ell}_{t-1})$. It follows from Equation (4.35) that, for every $\mathcal{A}$ and i^*, $\mathcal{R}_T(i^*, \mathcal{A})$ increases with T. Hence, without loss of generality we may assume that $T = \lfloor (\log M)/(16\Delta^2) \rfloor$. The maximum pseudo-regret of $\mathcal{A}$ on the instances $\mathbb{P}_{i^*}$ is lower-bounded as follows:

$$
\sup_{1\leqslant i^* \leqslant M} \mathcal{R}_T(i^*, \mathcal{A}) \geqslant \frac{1}{M} \sum_{1\leqslant i^* \leqslant M} \mathcal{R}_T(i^*, \mathcal{A}) = \frac{1}{M} \sum_{1\leqslant i^* \leqslant M} \Delta \sum_{t=1}^T \mathbb{E}_{i^*}[1 - v_{i^*,t}]. \tag{4.36}
$$

We now "randomize" the algorithm $\mathcal{A}$, by replacing it with a randomized algorithm which picks expert i at time t with probability $v_{i,t}$. Formally, let $\widetilde{P} = \mathcal{U}([0,1])^{\otimes T}$ be the distribution of T independent uniform random variables on $[0,1]$, and denote $\widetilde{\mathbb{P}}_{i^*} = \mathbb{P}_{i^*} \otimes \widetilde{P}$ for $i^* \in \{1, \ldots, M\}$. Furthermore, for every $\boldsymbol{v} \in \mathcal{P}_M$, let $I_{\boldsymbol{v}} : [0,1] \to \{1, \ldots, M\}$ be a measurable map such that $\mathbb{P}(I_{\boldsymbol{v}}(U) = i) = v_i$ for every $i \in \{1, \ldots, M\}$, where $U \sim \mathcal{U}([0,1])$. For every sequence of losses $\boldsymbol{\ell}_1, \ldots, \boldsymbol{\ell}_T$ and random variables $U_1, \ldots, U_T$ and every $1 \leqslant t \leqslant T$, let $I_t = I_{\boldsymbol{v}_t}(U_t)$, where $\boldsymbol{v}_t = A_t(\boldsymbol{\ell}_1, \ldots, \boldsymbol{\ell}_t)$.

Denote by $\widetilde{\mathbb{E}}_{i^*}$ the expectation with respect to $\widetilde{\mathbb{P}}_{i^*}$. By definition of $I_{\boldsymbol{v}}$, we have $\mathbb{E}_{i^*}[v_{i^*,t}] = \widetilde{\mathbb{E}}_{i^*}[\mathbf{1}(I_t = i^*)]$ so that, denoting $N_i = \sum_{i=1}^{T} \mathbf{1}(I_t = i)$ the number of times expert i is picked,

$$\sum_{t=1}^{T} \mathbb{E}_{i^*}[1 - v_{i^*,t}] = \widetilde{\mathbb{E}}_{i^*}[T - N_{i^*}] \geqslant \mathbb{P}_{i^*}(N_{i^*} \leqslant T/2) \cdot \frac{T}{2}.$$

Hence, letting $A_i \subseteq [0,1]^{M \times T} \times [0,1]^T$ be the event $\{N_i > T/2\}$, Equation (4.36) implies that

$$\sup_{1 \leqslant i^* \leqslant M} \mathcal{R}_T(i^*, \mathcal{A}) \geqslant \frac{\Delta T}{2} \times \frac{1}{M} \sum_{1 \leqslant i^* \leqslant M} \left(1 - \widetilde{\mathbb{P}}_{i^*}(A_{i^*})\right). \tag{4.37}$$

It now remains to upper bound $\frac{1}{M} \sum_{i^*} \widetilde{\mathbb{P}}_{i^*}(A_{i^*})$. To do this, first note that the events A_{i^*}, $1 \leqslant i^* \leqslant M$, are pairwise disjoint. Hence, Fano's inequality (see Gerchinovitz et al., 2017, p.2) implies that, for every distribution $\widetilde{\mathbb{Q}}$ on $[0,1]^{M \times T} \times [0,1]^T$,

$$\frac{1}{M} \sum_{1 \leqslant i^* \leqslant M} \widetilde{\mathbb{P}}_{i^*}(A_{i^*}) \leqslant \frac{1}{\log M} \left\{ \frac{1}{M} \sum_{1 \leqslant i^* \leqslant M} \mathrm{KL}(\widetilde{\mathbb{P}}_{i^*}, \widetilde{\mathbb{Q}}) + \log 2 \right\} \tag{4.38}$$

where $\mathrm{KL}(\mathbb{P}, \mathbb{Q})$ denotes the Kullback-Leibler divergence between $\mathbb{P}$ and $\mathbb{Q}$. Here, we take $\widetilde{\mathbb{Q}} = \mathbb{Q} \otimes \widetilde{P}$, where $\mathbb{Q}$ is the product of Bernoulli distributions $\mathcal{B}(1/2)^{\otimes T}$. This choice leads to

$$\mathrm{KL}(\widetilde{\mathbb{P}}_{i^*}, \widetilde{\mathbb{Q}}) = \mathrm{KL}(\mathbb{P}_{i^*}, \mathbb{Q}) = T \cdot \mathrm{KL}(\mathcal{B}(1/2 - \Delta), \mathcal{B}(1/2)) \leqslant 4T\Delta^2 \leqslant \frac{\log M}{4},$$

where the first bound is obtained by comparing KL and χ^2 divergences (Tsybakov, 2009, Lemma 2.7). Hence, inequality (4.38) becomes (recalling that $M \geqslant 4$)

$$\frac{1}{M} \sum_{1 \leqslant i^* \leqslant M} \widetilde{\mathbb{P}}_{i^*}(A_{i^*}) \leqslant \frac{(\log M)/4}{\log M} + \frac{\log 2}{\log M} \leqslant \frac{3}{4};$$

plugging this into (4.37) yields, noting that $T = \lfloor (\log M)/(16\Delta^2) \rfloor \geqslant (\log M)/(32\Delta^2)$ since $(\log M)/(16\Delta^2) \geqslant 1$ (as $M \geqslant 4$ and $\Delta \leqslant \frac{1}{4}$),

$$\sup_{1 \leqslant i^* \leqslant M} \mathcal{R}_T(i^*, \mathcal{A}) \geqslant \frac{\Delta T}{2} \times \frac{1}{4} \geqslant \frac{\log M}{256\Delta}.$$

This concludes the proof.

4.7.3 Proof of Theorem 4.2 and Corollary 4.1

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 4.2 implies

$$\begin{aligned}
R_T &\leqslant c_1\sqrt{T \log M} \\
&\leqslant c_1\sqrt{\tau_0 \log M} + c_1\sqrt{t_0 \log M} \\
&\leqslant c_1\sqrt{\tau_0 \log M} + c_1\sqrt{\log M} + \frac{\sqrt{8}\log(M/\Delta)}{c_0\Delta} \tag{4.39}
\end{aligned}$$

which implies (4.9) with $c_2 = c_1 + \frac{\sqrt{8}}{c_0}$ and $c_3 = \frac{\sqrt{8}}{c_0}$ (since $1 \leqslant \sqrt{\log M} \leqslant \frac{\log M}{\Delta}$). From now on, assume that $T \geqslant t_1 + 1$. Since $T \geqslant \tau_0$, we have $R_T = \widehat{L}_T - L_{i^*,T}$, so that

$$R_T = \widehat{L}_{t_1} - L_{i^*,t_1} + \sum_{t=t_1+1}^{T} \left(\widehat{\ell}_t - \ell_{i^*,t} \right). \tag{4.40}$$

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

$$\widehat{\ell}_t - \ell_{i^*,t} = \sum_{i \neq i^*} v_{i,t}(\ell_{i,t} - \ell_{i^*,t})$$

$$\leqslant \sum_{i \neq i^*} v_{i,t}$$

$$= \sum_{i \neq i^*} \frac{e^{-\eta_t(L_{i,t-1} - L_{i^*,t-1})}}{1 + \sum_{j \neq i^*} e^{-\eta_t(L_{j,t-1} - L_{i^*,t-1})}}$$

$$\leqslant \sum_{i \neq i^*} e^{-c_0 \sqrt{(\log M)/t} \times \Delta(t-1)} \tag{4.41}$$

$$\leqslant M e^{-c_0 \Delta \sqrt{(t-1)(\log M)/2}}$$

$$\leqslant \left(M e^{-c_0 \Delta \sqrt{t_0 (\log M)/8}} \right) e^{-c_0 \Delta \sqrt{(t-1)/8}} \tag{4.42}$$

$$\leqslant \Delta e^{-c_0 \Delta \sqrt{(t-1)/8}} \tag{4.43}$$

where (4.41) 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 $t - 1 \geqslant t_1 \geqslant \tau_0$), (4.42) from the fact that $t - 1 \geqslant t_0$ and $\log M \geqslant 1$, and (4.43) from the fact that $M e^{-c_0 \Delta \sqrt{t_0 (\log M)/8}} \leqslant \Delta$. Summing inequality (4.43), we obtain

$$\sum_{t=t_1+1}^{T} \left(\widehat{\ell}_t - \ell_{i^*,t} \right) \leqslant \sum_{t=t_1+1}^{T} \Delta e^{-c_0 \Delta \sqrt{(t-1)/8}}$$

$$\leqslant \Delta \sum_{t \geqslant 1} e^{-c_0 \Delta \sqrt{t/8}}$$

$$\leqslant \Delta \times \frac{2}{(c_0 \Delta / \sqrt{8})^2} \tag{4.44}$$

$$= \frac{16}{c_0^2 \Delta} \tag{4.45}$$

where (4.44) follows from Lemma 4.1. Combining (4.40), (4.39) and (4.45) proves Theorem 4.2 with $c_2 = c_1 + \frac{\sqrt{8}}{c_0}$, $c_3 = \frac{\sqrt{8}}{c_0}$ and $c_4 = \frac{16}{c_0^2}$.

Proof of Corollary 4.1. Define $\tau = \sup\{t \geqslant 0, \exists i \neq i^*, L_{i,t} - L_{i^*,t} \leqslant \frac{\Delta t}{2}\}$. By Lemma 4.2 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 4.2, this implies that, with probability at least $1 - \varepsilon$,

$$R_T \leqslant c_1 \sqrt{\tau \log M} + \frac{c_2 \log M + c_3 \log \Delta^{-1} + c_4}{\Delta/2}$$

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

where c_2, c_3, c_4 are the constants of Theorem 4.2. The bound (4.11) on the pseudo-regret is obtained similarly from Theorem 4.2, by using the fact that $\mathcal{R}_T \leqslant \mathbb{E}[R_T]$ and

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

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

Lemma 4.2. *Let* $(\ell_{i,t})_{1\leqslant i\leqslant M, t\geqslant 1}$ *be as in Theorem 4.1. Denote* $\tau = \sup\{t \geqslant 0, \exists i \neq i^*, L_{i,t} - L_{i^*,t} \leqslant \frac{\Delta t}{2}\}$. *We have*

$$\mathbb{E}[\tau] \leqslant 1 + \frac{8(\log M + 1)}{\Delta^2}, \tag{4.46}$$

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

$$\mathbb{P}\Big(\tau \geqslant \frac{8(\log M + \log \varepsilon^{-1})}{\Delta^2}\Big) \leqslant \varepsilon. \tag{4.47}$$

Proof of Lemma 4.2. For every $i \neq i^*$ and $t \geqslant 1$, let $\Delta_{i,t} := \mathbb{E}[\ell_{i,t} - \ell_{i^*,t}|\mathcal{F}_{t-1}]$. Using the Hoeffding-Azuma's maximal inequality 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 $\Delta_{i,t} - 1 \leqslant Z_{i,t} \leqslant \Delta_{i,t} + 1$), together with the fact that $\Delta_{i,t} \geqslant \Delta$, implies that

$$\mathbb{P}\left(\exists t \geqslant t_0, L_{i,t} - L_{i^*,t} \leqslant \frac{\Delta t}{2}\right) \leqslant \mathbb{P}\left(\sup_{t \geqslant t_0} \frac{1}{t}\left(\sum_{s=1}^{t} Z_{i,s}\right) \geqslant \frac{\Delta}{2}\right) \leqslant e^{-t_0 \Delta^2/8}. \tag{4.48}$$

By a union bound, equation (4.48) implies that

$$\mathbb{P}(\tau \geqslant t_0) \leqslant M e^{-t_0 \Delta^2/8}. \tag{4.49}$$

Solving for the probability level in (4.49) yields the high probability bound (4.47) on τ. The bound on τ in expectation (4.46) ensues by integrating the high-probability bound over ε. $\quad\square$

We recall Hoeffding-Azuma's maximal inequality for bounded martingale difference sequences (Hoeffding, 1963; Azuma, 1967). While it follows from a standard argument, we provide a short proof for completeness, since the inequality given in Proposition 4.5 below differs slightly from the one given in Hoeffding (1963).

Proposition 4.5 (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* $a \geqslant 0$:

$$\mathbb{P}\left(\sup_{m \geqslant n} \frac{S_m}{m} \geqslant a\right) \leqslant e^{-na^2/2}. \tag{4.50}$$

Proof. Fix $\lambda > 0$. By Hoeffding's inequality, $\mathbb{E}[e^{\lambda Z_t}|\mathcal{F}_{t-1}] \leqslant e^{\lambda^2/2}$, so that the sequence $M_t^\lambda := \exp\left(\lambda S_t - \lambda^2 t/2\right)$ is a positive supermartingale. Hence, Doob's supermartingale inequality implies that for $\varepsilon \in (0,1]$:

$$\mathbb{P}\Big(\sup_{t \geqslant 1} M_t^\lambda \geqslant \frac{1}{\varepsilon}\Big) \leqslant \frac{\mathbb{E}[M_0^\lambda]}{1/\varepsilon} = \varepsilon. \tag{4.51}$$

Rearranging (4.51) and letting $\lambda = \sqrt{2\log(1/\varepsilon)/n}$ yields: with probability $1 - \varepsilon$, for every $t \geqslant n$,

$$\frac{S_t}{t} \leqslant \frac{\log(1/\varepsilon)}{\lambda t} + \frac{\lambda}{2} = \sqrt{\frac{\log(1/\varepsilon)}{2}}\left(\frac{\sqrt{n}}{t} + \frac{1}{\sqrt{t}}\right) \leqslant \sqrt{\frac{2\log(1/\varepsilon)}{n}} \ . \tag{4.52}$$

Setting $\varepsilon = e^{-na^2/2}$ in (4.52) gives the desired bound. $\qquad\square$

4.7.4 Proof of Proposition 4.3

Note that, since the loss vectors ℓ_t are in fact deterministic, $\mathcal{R}_T = R_T$. 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

$$\begin{aligned}
R_T &= \sum_{t=1}^{T}\sum_{i=2}^{M} v_{i,t}(\ell_{i,t} - \ell_{1,t}) \\
&= \sum_{t=1}^{T}\sum_{i=2}^{M} \frac{\exp\left(-\frac{c}{\sqrt{T}}(L_{i,t-1} - L_{1,t-1})\right)}{1 + \sum_{2\leqslant i'\leqslant M}\exp\left(-\frac{c}{\sqrt{T}}(L_{i',t-1} - L_{1,t-1})\right)} \\
&= \sum_{t=1}^{T} \frac{(M-1)\exp\left(-\frac{c}{\sqrt{T}}(t-1)\right)}{1 + (M-1)\exp\left(-\frac{c}{\sqrt{T}}(t-1)\right)} \ .
\end{aligned} \tag{4.53}$$

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 (4.53) that

$$R_T \geqslant \sum_{t=1}^{T\wedge(t_0+1)} \frac{(M-1)\exp\left(-\frac{c}{\sqrt{T}}(t-1)\right)}{1 + (M-1)\exp\left(-\frac{c}{\sqrt{T}}(t-1)\right)} \geqslant \frac{1}{3}\min(T, t_0+1) \tag{4.54}$$

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 (4.13), it remains to note that

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

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 (4.13) applies; further bounding $T_{k-1} \geqslant \frac{T}{4}$ establishes (4.14).

4.7.5   Proof of Proposition 4.4

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

$$\mathbb{E}[(\widehat{\ell}_t - \ell_{i^*,t})^2] \leqslant \mathbb{E}\left[\sum_{i=1}^{M} v_{i,t}(\ell_{i,t} - \ell_{i^*,t})^2\right] \tag{4.55}$$

$$= \mathbb{E}\left[\sum_{i=1}^{M} v_{i,t}\mathbb{E}\left[(\ell_{i,t} - \ell_{i^*,t})^2 | \mathcal{F}_{t-1}\right]\right]$$

$$\leqslant B\mathbb{E}\left[\sum_{i=1}^{M} v_{i,t}\mathbb{E}\left[\ell_{i,t} - \ell_{i^*,t} | \mathcal{F}_{t-1}\right]^\beta\right] \tag{4.56}$$

$$\leqslant B\mathbb{E}\left[\sum_{i=1}^{M} v_{i,t}\mathbb{E}\left[\ell_{i,t} - \ell_{i^*,t} | \mathcal{F}_{t-1}\right]\right]^\beta \tag{4.57}$$

$$= B\mathbb{E}[\widehat{\ell}_t - \ell_{i^*,t}]^\beta \tag{4.58}$$

where inequalities (4.55) and (4.57) come from Jensen's inequality, and (4.56) from the Bernstein condition (4.15). Taking the expectation of the regret bound (4.16), we obtain

$$\mathbb{E}[R_{i^*,T}] \leqslant \mathbb{E}\left[C_1\sqrt{(\log M)\sum_{t=1}^{T}(\widehat{\ell}_t - \ell_{i^*,t})^2} + C_2\log M\right]$$

$$\leqslant C_1\sqrt{(\log M)\sum_{t=1}^{T}\mathbb{E}\left[(\widehat{\ell}_t - \ell_{i^*,t})^2\right]} + C_2\log M \tag{4.59}$$

$$\leqslant C_1\sqrt{(\log M)B\sum_{t=1}^{T}\mathbb{E}\left[\widehat{\ell}_t - \ell_{i^*,t}\right]^\beta} + C_2\log M$$

$$= C_1\sqrt{BT\log M}\left(\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\left[\widehat{\ell}_t - \ell_{i^*,t}\right]^\beta\right)^{1/2} + C_2\log M$$

$$\leqslant C_1\sqrt{BT\log M}\left(\frac{\mathbb{E}[R_{i^*,T}]}{T}\right)^{\beta/2} + C_2\log M \tag{4.60}$$

where inequalities (4.59) and (4.60) come from Jensen's inequality. Letting $r = \mathbb{E}[R_{i^*,T}]/T$ and $u = (\log M)/T$, inequality (4.60) writes $r \leqslant C_1\sqrt{Bu}\,r^{\beta/2} + C_2 u$. This implies that (depending on which of these two terms is larger) either $r \leqslant 2C_2 u$, 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

$$r \leqslant (2C_1)^{\frac{2}{2-\beta}}(Bu)^{\frac{1}{2-\beta}} + 2C_2 u\,,$$

which amounts to

$$\mathbb{E}[R_{i^*,T}] \leqslant C_3(B\log M)^{\frac{1}{2-\beta}}T^{\frac{1-\beta}{2-\beta}} + C_4\log M \tag{4.61}$$

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

4.7.6 Proof of Theorem 4.3

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

$$
\begin{aligned}
\mathcal{R}_T &= \sum_{t=1}^{T}\sum_{i\neq 1}\mathbb{E}[v_{i,t}(\ell_{i,t} - \ell_{1,t})] \\
&= \Delta\sum_{t=1}^{T}\frac{(M-1)e^{-\eta_t\Delta(t-1)}}{1 + (M-1)e^{-\eta_t\Delta(t-1)}} \\
&\geqslant \frac{\Delta}{3}\sum_{t=1}^{T}\mathbf{1}\Big((M-1)e^{-\eta_t\Delta(t-1)}\geqslant \frac{1}{2}\Big) \\
&\geqslant \frac{\Delta}{3}\sum_{t=1}^{T}\mathbf{1}\Big(Me^{-c_0\Delta\sqrt{(t-1)\log M}}\geqslant 1\Big) & (4.62) \\
&\geqslant \frac{\Delta}{3}\times\min\Big(\frac{\log M}{c_0^2\Delta^2},T\Big) \\
&= \frac{1}{3}\min\Big(\frac{1}{c_0}\sqrt{T\log M},T\Big), & (4.63)
\end{aligned}
$$

where (4.62) relies on the inequalities $2(M-1)\geqslant M$ and $(t-1)/\sqrt{t}\leqslant\sqrt{t-1}$ for $M\geqslant 2, t\geqslant 1$, while (4.63) 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.

4.7.7 Proof of Theorem 4.4

Assume that the loss vectors $\boldsymbol{\ell}_1,\boldsymbol{\ell}_2,\ldots$ are i.i.d., and denote $i^* = \arg\min_{1\leqslant i\leqslant M}\mathbb{E}[\ell_{i,t}]$ (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 $j\in\{1,\ldots,M\}$ such that $\Delta_j = \Delta$. The Decreasing Hedge algorithm with learning rate $\eta_t = c_0\sqrt{(\log M)/t}$ satisfies

$$
\begin{aligned}
\mathcal{R}_T &= \sum_{t=1}^{T}\sum_{i\neq i^*}\mathbb{E}[v_{i,t}]\Delta_i \\
&\geqslant \Delta\sum_{t=1}^{T}\mathbb{E}\left[\frac{\sum_{i\neq i^*}e^{-\eta_t(L_{i,t-1}-L_{i^*,t-1})}}{1 + \sum_{i\neq i^*}e^{-\eta_t(L_{i,t-1}-L_{i^*,t-1})}}\right] \\
&\geqslant \Delta\sum_{t=1}^{T}\mathbb{E}\left[\frac{e^{-\eta_t(L_{j,t-1}-L_{i^*,t-1})}}{1 + e^{-\eta_t(L_{j,t-1}-L_{i^*,t-1})}}\right] & (4.64) \\
&\geqslant \frac{\Delta}{3}\sum_{t=1}^{T}\mathbb{E}\left[\mathbf{1}\Big(e^{-\eta_t(L_{j,t-1}-L_{i^*,t-1})}\geqslant\frac{1}{2}\Big)\right] \\
&= \frac{\Delta}{3}\sum_{t=1}^{T}\mathbb{P}\big(\eta_t(L_{j,t-1}-L_{i^*,t-1})\leqslant\log 2\big) & (4.65)
\end{aligned}
$$

195

where (4.64) 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 $1 \leqslant t \leqslant 1 + \frac{a^2}{4\Delta^2}$:

$$\mathbb{P}\left(\eta_t(L_{j,t-1} - L_{i^*,t-1}) > \log 2\right) = \mathbb{P}\left(L_{j,t-1} - L_{i^*,t-1} - \Delta(t-1) > a\sqrt{t} - \Delta(t-1)\right)$$

$$\leqslant \mathbb{P}\left(L_{j,t-1} - L_{i^*,t-1} - \Delta(t-1) > \frac{a\sqrt{t-1}}{2}\right) \tag{4.66}$$

$$\leqslant e^{-a^2/8} \tag{4.67}$$

where inequality (4.66) 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 (4.67) 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 (4.65) and (4.67) 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 $T \geqslant \frac{1}{4\Delta^2} \geqslant \frac{a^2}{4\Delta^2}$:

$$\mathbb{E}\left[R_T\right] \geqslant \frac{\Delta}{3} \min\left(\frac{a^2}{4\Delta^2}, T\right)(1 - e^{-1/8})a^2 = \frac{(1 - e^{-1/8})a^4}{12\Delta} \geqslant \frac{1}{450c_0^4(\log M)^2\Delta}, \tag{4.68}$$

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