JMLR: Workshop and Conference Proceedings vol 30 (2013) 1–21

Regret Minimization for Branching Experts

Eyal Gofer EYALGOFE@POST.TAU.AC.IL
Tel Aviv University, Tel Aviv, Israel

Nicolò Cesa-Bianchi NICOLO.CESA-BIANCHI@UNIMI.IT
Università degli Studi di Milano, Milan, Italy

Claudio Gentile CLAUDIO.GENTILE@UNINSUBRIA.IT
Università degli Studi dell'Insubria, Varese, Italy

Yishay Mansour MANSOUR@POST.TAU.AC.IL
Tel Aviv University, Tel Aviv, Israel

Abstract

We study regret minimization bounds in which the dependence on the number of experts is replaced by measures of the realized complexity of the expert class. The measures we consider are defined in retrospect given the realized losses. We concentrate on two interesting cases. In the first, our measure of complexity is the number of different "leading experts", namely, experts that were best at some point in time. We derive regret bounds that depend only on this measure, independent of the total number of experts. We also consider a case where all experts remain grouped in just a few clusters in terms of their realized cumulative losses. Here too, our regret bounds depend only on the number of clusters determined in retrospect, which serves as a measure of complexity. Our results are obtained as special cases of a more general analysis for a setting of branching experts, where the set of experts may grow over time according to a tree-like structure, determined by an adversary. For this setting of branching experts, we give algorithms and analysis that cover both the full information and the bandit scenarios.

Keywords: Regret Minimization, Hedge Algorithm, Structured Experts.

1. Introduction

Prediction with expert advice (Vovk, 1990; Littlestone and Warmuth, 1994) is a crisp abstract framework for studying sequential decision problems in nonstochastic settings. In this paper, we focus on the following special case —also known as decision-theoretic online learning— of the general experts framework. There are N experts (or, equivalently, actions). At each time step $t = 1, 2, \ldots$ a learner selects a distribution $\boldsymbol{p}_t$ over the experts and, simultaneously, an adversary reveals a vector $\boldsymbol{\ell}_t = (\ell_{1,t}, \ldots, \ell_{N,t}) \in [0,1]^N$ of expert losses. The learner then incurs a loss of $\widehat{\ell}_t = \boldsymbol{p}_t \cdot \boldsymbol{\ell}_t$. We use $L_{i,t} = \sum_{s=1}^{t} \ell_{i,s}$ and $\widehat{L}_t = \sum_{s=1}^{t} \widehat{\ell}_s$ to denote the total loss of expert i after t prediction steps and the total loss of the learner after t prediction steps, respectively. The goal of the learner in this repeated game is to control the regret $R_T = \widehat{L}_T - L_T^*$ over T, irrespective of the adversary's choices of $\boldsymbol{\ell}_t$, where $L_T^* = \min_{j=1,\ldots,N} L_{j,T}$ is the total loss of the best expert after T steps.

The distribution chosen by the learner is naturally interpreted as a random choice over the experts. Crucially, the learner makes this choice given full information of past losses incurred by every expert. An important variation on this full information setting limits

the learner's observations in every round of play only to the loss incurred by the expert it chose. This is known as the adversarial multi-armed bandit (or simply bandit) setting —see (Bubeck and Cesa-Bianchi, 2012) for a recent survey.

In these settings, a great deal of attention has been devoted to controlling the dependence of R_T on the time horizon T. The well-known Hedge algorithm of Freund and Schapire (1995) offers the most basic form of regret control in the full information case. Hedge has a single parameter $\eta \geq 0$. At time t, the algorithm maintains a weight $w_{i,t} > 0$ for each expert $i = 1, \ldots, N$, and uses the distribution $\boldsymbol{p}_t$ defined by $p_{i,t} = w_{i,t}/W_t$, where $W_t = \sum_{i=1}^{N} w_{i,t}$. The weights decrease exponentially in the experts' losses: $w_{i,t} = e^{-\eta L_{i,t-1}}$, where $L_{i,0} = 0$. If η is tuned solely as a function of time, we achieve the so-called zero-order regret bounds, which have the form $\mathcal{O}(\sqrt{T \ln N})$. More refined bounds, called first-order bounds, are obtained when η is allowed to depend on the performance L_T^* of the best expert. These bounds take the form $\mathcal{O}(\sqrt{L_T^* \ln N})$. More recent second-order bounds (Cesa-Bianchi et al., 2007; Hazan and Kale, 2008; Chiang et al., 2012) replace the dependence on L_T^* by quantities that measure the variability of the sequence of loss vectors $\boldsymbol{\ell}_t$. In the bandit setting regret bounds followed a similar evolution, starting from the zero-order bound $\mathcal{O}\big(\sqrt{TN \ln N}\big)$ on the expected regret attained by the randomized Exp3 algorithm of Auer et al. (2002).

This thread of research has not only produced better regret bounds and new algorithmic techniques, but has also allowed the introduction of new and interesting applications. For example, second-order bounds can be related to volatility, which is used in pricing derivatives (DeMarzo et al., 2006; Gofer and Mansour, 2011). Yet, the progress in improving the analysis with respect to the number N of experts has so far been more limited. In this paper, we design prediction algorithms with refined bounds in terms of N. Our algorithms are able to control regret when N may grow over time, but the experts' cumulative losses enjoy specific patterns that naturally occur in practical scenarios. These patterns are captured by the notion of "branching experts", in which the addition of new experts to the pool creates a tree structure. Although our algorithms are designed for this branching experts setting, when applied to the standard N-expert setting they deliver regret bounds where the "complexity" term $\log N$ associated with the set of N experts does not occur.

For the sake of illustration, consider the following sequential path-planning problem on a graph (Kalai and Vempala, 2005). The number N of available paths from source to destination (i.e., experts in the game) is very large, and the loss of a path picked at time t is the sum over the current costs (at the same time t) of the edges included in the path. One may expect that in T prediction steps only a small number Λ_T of paths have become "leaders" at any time $t \leq T$, where a path i is leader at time t if its cumulative loss $L_{i,t}$ is the smallest over all paths. We show that a variant of Hedge, played over a growing pool of leaders,[1] achieves a regret bound that only depends on the number of leaders, rather than on the number of experts. In general, in a game with N experts and Λ_T leaders, we prove that the regret of our modified Hedge is $\mathcal{O}\big(\Lambda_T(1 + \ln L_T^*) + \sqrt{L_T^* \Lambda_T}\big)$, independent of N. Our result is actually phrased in a more general model, where the notion of leader is parameterized by a value $\alpha \geq 0$ and Λ_T is replaced by $\Lambda_{\alpha,T}$. This value represents the edge an expert must have over the loss of the previous leaders in order to become leader itself.

1. Note that in this example a leader can be found efficiently by solving a shortest path problem.

A second natural scenario we consider is one where the cumulative losses of all experts remain clustered around the loss values of a few experts. Intuitively, as far as regret is concerned experts that have similar cumulative losses are interchangeable. Hence, working with one representative in each cluster is a convenient approximation of the original problem, and one expects the regret to be controlled by the number of clusters, rather than by the overall number of experts. As before, we make a reduction to the setting of a growing pool of experts. We start with a single cluster (all experts start off with zero loss) and then split a cluster (i.e., increase the pool of experts by at least one) whenever the largest cumulative loss difference within the cluster exceeds some threshold value $\alpha \geq 0$. We prove that the regret of our Hedge variant is at most of order $\mathcal{N}_{\alpha,T}(1+\alpha\mathcal{N}_{\alpha,T})(1+\ln L_T^*) + \sqrt{L_T^*\mathcal{N}_{\alpha,T}}$, where $\mathcal{N}_{\alpha,T}$ is the number of α-clusters after T steps.

In both of the above settings, of few leading experts and of clustered experts, our algorithm is essentially optimal: We prove that the main terms of both regret bounds, $\sqrt{L_T^*\Lambda_{\alpha,T}}$ and $\sqrt{L_T^*\mathcal{N}_{\alpha,T}}$, are only improvable by constant factors. The same result is also proven for the general case of a growing set of experts.

We then turn our attention to the bandit setting. Here, as a motivating scenario for the growing set of experts, consider a framework where we apply heuristics to solve a sequence of instances of a hard optimization problem. This task can be naturally cast in a bandit setting, where just a single heuristic is tested on each instance. Now suppose that new heuristics become available as time goes by, and we add them to the pool of candidate heuristics in order to improve our chances. Some of them might be variants of other heuristics in the pool, and some others might be completely new. In all cases, we would like to control the regret of the bandit algorithm with respect to the growing pool of heuristics. Now, if a variant i' of some heuristic i already in the pool becomes available at time t, then it is reasonable to compute the regret against a pair of compound experts that use i up to time t, and from then on either i or its variant i'. On the other hand, if a heuristic k unrelated to any other in the pool is added at time t, then we want to compare to a pair of compound experts that use the best heuristic up to time t and then either stick to it, or switch to k.

In this context, we introduce a new nontrivial modification of Exp3, and show that its expected regret is at most of order $\big(1+(\ln f)/(\ln N_T)\big)\sqrt{TN_T\ln N_T}$, where N_T is the final number of experts, and f stands for the product of the degrees of nodes along the branch leading to the best expert. This factor may be as small as $\Theta(1)$ and is always bounded by K^{d_T}, where d_T is the number of time steps in which some new expert was added to the pool, and K is the branching factor of the tree of experts.

2. Branching experts with full information

We consider a modification of prediction with expert advice where the adversary gradually reveals the experts to the learner. Specifically, the game between adversary and learner starts with one known expert. As the game progresses, the adversary may choose to reveal the existence of more experts. Once an expert is revealed, the learner may start using it by placing some weight on its decisions. The regret of a learner at the end of the game is measured w.r.t. all revealed experts.

Each newly revealed expert is given some history, in the form of a starting loss closely related to the cumulative loss of one of the previously revealed experts. In our setup we therefore consider each newly revealed expert as an approximate clone of an existing expert

in terms of its cumulative loss. From that point on, the new expert is allowed to freely diverge from its parent. Let $1, \ldots, N_t$ be the indices of the experts revealed after the first t rounds, with $N_0 = 1$. This process of approximate cloning results in a tree structure where, at any time $t \geq 1$, the root is the initial expert (the one at time $t = 0$) and the leaves correspond to the N_t experts participating in the game at time t. For convenience, when an expert is cloned, one of the clones is used to represent the continuation of the original expert (i.e., it is a perfect clone). We now describe the game in more detail.

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

1. For each expert $i = 1, \ldots, N_{t-1}$, the adversary reveals a set $C(i,t)$ of experts containing i itself and possibly additional approximate clones. The adversary also reveals the past losses $L_{j,t-2}, \ell_{j,t-1}$ for each $j \in C(i,t)$. The new experts are indexed by $N_{t-1}+1, \ldots, N_t$.
2. The learner chooses $\boldsymbol{p}_t = \big(p_{1,t}, \ldots, p_{N_t,t}\big)$.
3. The adversary reveals losses $\boldsymbol{\ell}_t = \big(\ell_{1,t}, \ldots, \ell_{N_t,t}\big)$, and the learner suffers a loss $\widehat{\ell}_t = \boldsymbol{p}_t \cdot \boldsymbol{\ell}_t$.

The reason why the pair $L_{j,t-2}, \ell_{j,t-1}$ is revealed instead of its sum $L_{j,t-2} + \ell_{j,t-1}$ is explained in Footnote 4. We use $m(t) = \operatorname{argmin}_{i=1,\ldots,N_t} L_{i,t}$ to denote the index of the best expert in the first t steps (where we take the smallest such index in case of a tie), so that $L_T^* = L_{m(T),T}$. The regret is then defined as usual, $R_T = \widehat{L}_T - L_T^*$. In order to have the standard expert scenario as a special case of the branching one, we set $L_{j,t}, \ell_{j,t} = 0$ for $t \leq 0$ and for all j. In the standard expert scenario $N_1 = |C(1,1)|$ corresponds to the number N of experts. In order to facilitate a comparison to the standard expert bounds, we express our branching experts bounds in terms of this quantity N_1.

In order to elucidate the combinatorial tree structure underlying this notion of regret, consider the simple case when $L_{j,t-1} = L_{i,t-1}$ for all $j \in C(i,t)$, all $i = 1, \ldots, N_{t-1}$, and all t. Hence, each new expert j starts off as a perfect clone of its parent expert i. Now, at each time T the learner is competing with the set of N_T compound experts associated with the paths $i_0, i_1, \ldots, i_T$, where i_0 is the "root expert" and $i_t \in C(i_{t-1}, t)$ for all $t = 1, \ldots, T$. This combinatorial interpretation will help in comparing our results to other settings of compound experts.

When clones are not perfect, we use $\alpha_{i,t-1} = \max\big\{\big|L_{j,t-2} - L_{k,t-2}\big| : j,k \in C(i,t)\big\}$ to denote the "diameter" of a split.[2] Since the learner has no control over $\alpha_{i,t}$ at each split, he may suffer the sum of those elements from the root to the leaf of the best expert in the form of regret. Furthermore, the splits increase the number of experts, which means that upper bounds on the regret must also grow. Interestingly, there are natural scenarios where the above quantities are small. In these scenarios, described in Section 5, the role of revealing new experts is taken from the adversary and *given to the learner*. Specifically, all the experts are known at the beginning of the game, and it is the learner who decides to gradually start using some of them. Generally speaking, in such scenarios the vast majority

2. Note that perfect cloning is not equivalent to a split with diameter zero. This discrepancy is due to technical issues that arise in Section 5.

of experts are either poor candidates for being the best expert, or perform comparably to some other experts, making them safe for the learner to ignore. The important observation to be made is that any algorithm and analysis that hold when the splits are determined by the adversary may also be used if the splits are determined by the learner. Although referring to the best overall expert rather than the best revealed expert clearly increases the regret, this difference will be bounded for the scenarios we consider, and easily taken into account.

3. Related work

Prediction with expert advice when the pool of experts varies over time has been investigated in the past. The model of "sleeping experts", or "specialists" (Freund et al., 1997; Blum and Mansour, 2007), allows a subset A_t of the experts to be awake at each time step t, where A_t is determined by a time-selection function. The regret is then measured against each expert only on the time steps when it was awake. Although we may view our setting of branching experts as a special case of sleeping experts (where experts are progressively woken up), their notion of regret is different from ours, since we compete against *compound* experts, associated with paths on the expert tree. Therefore, the two settings are incomparable.

Branching experts can also be viewed as special cases of more general combinatorial constructions, like the permutation experts of Kleinberg et al. (2010) or the shifting experts of Herbster and Warmuth (1998). Although in some cases these more general settings can be extended to accommodate growing sets of experts —see, e.g., (Shalizi et al., 2011), the resulting regret bounds are much worse than ours, mainly due to the dependence on $\ln(N_T T)$, where N_T is the total number of experts in the pool after T steps. Our bounds, instead, depend in a more detailed way on the structure of the expert tree, and replace $\ln(N_T T)$ with $\ln \Pi$, where Π depends on the splits on the branches that the leading experts belong to, and can thus be much smaller than N_T —see discussion in the next paragraph.

In the case of perfect cloning, which was mentioned in Section 2, Hedge may directly simulate the branching experts if we limit the number of rounds in which splits occur to at most d, and the maximal degree of a split to at most K, where d and K are preliminarily available. Taking $N_1 = 1$, this limits the final number of compound experts to at most K^d. An application of standard Hedge then leads to the regret bound $d \ln K + \sqrt{2 L_T^* d \ln K}$. Our modified Hedge, called Hed_C, virtually obtains the same bound (with slightly worse constants) without preliminary knowledge of the tree structure. However, we may do much better in cases where there are few splits along the branches where the set of leading experts occur. In fact, the main term in the regret bound of Hed_C is of order $\sqrt{L_T^* \ln \Pi}$, where Π can be $\Theta(1)$ even when N_T is exponential in T.

It is also instructive to consider a type of scenario in which any advantage of Hed_C over Hedge is removed. Specifically, assume that all the splits are announced in the initial rounds of the game. In these initial rounds no losses are incurred, and afterwards, the game proceeds with a fixed number of experts. It is possible to design such a scenario with $N+1$ experts (for a large enough N) such that Hed_C has regret $R_T = \Omega\big(\sqrt{L_T^* N}\big)$ while Hedge has the usual regret $R_T = \mathcal{O}\big(\ln N + \sqrt{L_T^* \ln N}\big)$ (see Claim 13 in the appendix). Since Theorem 3 in Section 4 gives an $\mathcal{O}\big(N + \sqrt{L_T^* N}\big)$ regret for this case, it is arguably a worst-case scenario for Hed_C. Clearly enough, in order to make Hed_C achieve a similar

regret as Hedge in this hard case, without losing much of its advantage in the others, it suffices to add a meta-Hedge algorithm aggregating the two.

In the bandit setting, our modification of Exp3 for branching experts, called Exp3.C, has an expected regret bound of the order $\big(1+(\ln f)/(\ln N_T)\big)\sqrt{TN_T \ln N_T}$, containing the factor $\sqrt{N_T}$ (see Section 7). It seems plausible that the modification of Exp3 for shifting experts, Exp3.S, could be extended to a growing pool of experts along the lines of (Shalizi et al., 2011). The resulting regret bound would be of the order $\sqrt{TN_T S_T \ln(N_T T)}$, where S_T is the number of times $i_t \neq i_{t+1}$ in the path $i_0, i_1, \ldots, i_T$ from the root to the best action $i_T = m(T)$. These two bounds appear to be incomparable.

We now move on to discussing related work in the standard N-expert setting, with fixed N. There are a few examples of expert algorithms whose regret bound does not depend on the number of experts. The most trivial example is Follow the Leader (FTL), namely, the algorithm that deterministically picks the current best expert. It is easy to see that the regret of FTL is bounded by the number of times the leader changes, *no matter if the same few leaders* keep alternating. Another trivial example is Hedge run with uneven initial weights, which implicitly assumes using a prior that peaks on a small set containing the best expert. A substantially less trivial example is the NormalHedge algorithm of Chaudhuri et al. (2009), and its refinement due to Chernov and Vovk (2010). Except for constant factors, a bound on the cumulative loss of these algorithms can be written as $\inf_{0\leq\epsilon\leq 1}\big(L_{\sigma(\epsilon N),T} + \sqrt{T\ln(1/\epsilon)}\big)$ where $\sigma(1),\ldots,\sigma(N)$ is a permutation of expert indices such that $L_{\sigma(1),T} \leq \cdots \leq L_{\sigma(N),T}$. This bound is incomparable to our bounds for few leaders and clustered experts. Indeed, it is easy to find examples where our bounds dominate NormalHedge's. (Think of one expert with constant total loss, and the remaining $N-1$ experts with linear losses always clustered around a single value: Then ϵ must be $1/N$ to ensure sublinear regret, while Λ_T and $\mathcal{N}_{\alpha,T}$ can be made constant.)

Finally, we could apply the Fixed Share algorithm of Herbster and Warmuth (1998) for shifting experts to the few leaders setting. By doing so, we would get a regret bound of the order of $\sqrt{TS_T \ln(NT)}$ where S_T is the number of times the current leader changes. Now, even in the best case for Fixed Share (i.e., when $S_T = \Lambda_T - 1$) our bound for Hed_C is still better by a factor of at least $\sqrt{\ln(NT)}$. Using more sophisticated shifting algorithms, like Mixing Past Posteriors (Bousquet and Warmuth, 2002), may improve on the $S_T \ln N$ term in the Fixed Share bound, but it does not affect the other term $S_T \ln T$. Recently, Koolen et al. (2012) gave a surprising Bayesian interpretation to Mixing Past Posteriors. Finding a similarly efficient Bayesian formulation of our branching experts construction is an open problem.

The few leaders and the clustered experts settings take advantage of specific "suboptimalities" in the loss sequence chosen by the adversary. Adaptive procedures able to take advantage of such suboptimalities were also proposed by Rakhlin et al. (2012). Relating those results to the ones derived in this paper remains an interesting open problem.

4. Adapting Hedge for the branching setup

The main change that is required to handle the new setup is deciding on weights for newly revealed experts. We will handle this problem by applying a general mechanism we term a *partial restart*. A partial restart redistributes the weights $w_{i,t}$ of existing experts among all

experts, old and new, without changing the sum of the weights. Following this redistribution, the usual exponential update step takes place.

Hedge with partial restarts
For each round $t = 1, 2, \ldots, T$

1. Get the new experts $N_{t-1}+1, \ldots, N_t$ together with their past losses $L_{j,t-2}, \ell_{j,t-1}$ for $j = N_{t-1}+1, \ldots, N_t$.
2. From $w_{1,t-1}, \ldots, w_{N_{t-1},t-1}$ compute new weights $w'_{1,t-1}, \ldots, w'_{N_t,t-1} \geq 0$ such that $W'_{t-1} = W_{t-1}$, where $W'_{t-1} = \sum_{i=1}^{N_t} w'_{i,t-1}$.
3. Update the new weights: $w_{i,t} = w'_{i,t-1} e^{-\eta \ell_{i,t-1}}$, and set $p_{i,t} = w_{i,t}/W_t$, for each $i = 1, \ldots, N_t$.
4. Observe losses $\boldsymbol{\ell}_t = \big(\ell_{1,t}, \ldots, \ell_{N_t,t}\big)$ and suffer loss $\widehat{\ell}_t = \boldsymbol{p}_t \cdot \boldsymbol{\ell}_t$.

Note that an ordinary (full) restart is equivalent to a partial restart where all experts are assigned equal weights. A partial restart, in contrast to a full restart, may preserve more information about the preceding run, depending on how $w'_{i,t}$ are defined.

If an algorithm is an augmentation of Hedge with partial restarts, its loss is upper bounded by an expression that depends explicitly on the number of restarts, and only implicitly on their exact nature. This bound is given in the next lemma.

Lemma 1 *Let $0 < \eta \leq 1$ and let A be an augmentation of Hedge with at most n partial restarts. Then* $\ln(W_{T+1}/W_1) \leq -(1 - e^{-\eta})(\widehat{L}_T - n)$.

We now describe a specific way of defining the weights $w'_{i,t}$ out of the weights $w_{i,t}$, and call Hed_C the resulting partially restarting variant of Hedge. The algorithm starts by setting $w_{j,1} = 1$ for every expert $j = 1, \ldots, N_1$, where $N_1 = |C(1,1)|$.[3] At time $t+1$, its partial restart stage distributes the weight of a parent expert i among experts in $C(i, t+1)$ while maintaining the same weight proportions as ordinary Hedge. Namely,

$$w'_{j,t} = w_{i,t} \frac{e^{-\eta L_{j,t-1}}}{\sum_{k \in C(i,t+1)} e^{-\eta L_{k,t-1}}}, \qquad \text{for every } j \in C(i, t+1)$$

(recall that $L_{j,t-1}$, for $j \in C(i,t+1)$, are all revealed to the algorithm).[4] Note that when $|C(i,t+1)| = 1$ —that is, expert i is not cloned at time $t+1$, then $w'_{i,t} = w_{i,t}$. As a step towards bounding the regret of Hed_C, we first lower bound the weights it assigns to experts.

Lemma 2 *If i is an expert at time $t \geq 1$, and $i_0, \ldots, i_t = i$ are the experts along the path from the root to i in the branching experts tree, then*

$$w_{i_t,t} \geq \exp\left(-\eta L_{i_t,t-1} - \eta \sum_{\tau=1}^{t-1} \alpha_{i_\tau,\tau}\right) \prod_{\tau=1}^{t-1} \max\Big\{1, 2\big|C(i_\tau, \tau+1)\big| - 2\Big\}^{-1} .$$

3. Alternatively, we may set $w_{1,0} = 1$ for the single expert existing at time $t = 0$. This causes a minor difference in the regret bounds.
4. The two-stage update $w_{i,t} \to w'_{i,t} \to w_{i,t+1}$ is the reason why pairs of losses $L_{j,t-1}$ and $\ell_{j,t}$ are provided for each new expert $i = N_t + 1, \ldots, N_{t+1}$.

We can now combine Lemmas 1 and 2 to upper bound the regret of Hed_C. In what follows we denote $\Pi_t = N_1 \prod_{\tau=1}^{t-1} \max\Big\{1, 2\big|C(i_\tau, \tau+1)\big| - 2\Big\}$ and $\mathcal{A}_t = \sum_{\tau=1}^{t-1} \alpha_{i_\tau,\tau}$, where $i_0, \ldots, i_t = m(t)$ is a path from the root to the best expert at time t. Note that neither Π_t nor $\mathcal{A}_t$ are monotone in t because the index of the best expert $m(t)$ changes over time. We also denote d_t for the number of rounds in $2, \ldots, t$ in which splits occurred.

Theorem 3 *Let $\mathcal{L}'$ and Π' be known upper bounds on $L_T^* + \mathcal{A}_{T+1}$ and Π_{T+1}, respectively. If $\eta = \ln\big(1 + \sqrt{(2\ln\Pi')/\mathcal{L}'}\big)$, then the regret R_T of Hed_C satisfies*

$$R_T \le d_T + \mathcal{A}_{T+1} + \ln\Pi' + \sqrt{2\mathcal{L}'\ln\Pi'} \ .$$

In the standard N-expert setting, $\alpha_{i,t} = 0$ for all i and t, implying $\mathcal{A}_{T+1} = 0$. Moreover, $d_T = 0$ and $\Pi_{T+1} = N$. Therefore, in this special case Theorem 3 recovers the standard regret bound of Hedge.

The requirement in Theorem 3 that bounds be known in advance may be relaxed by using a doubling trick. Since the index of the best expert $m(t)$ changes over time, the doubling is applied to a bound $\mathcal{L}'_\tau$ on the quantity $\max_{t\le\tau}\{L_t^* + \mathcal{A}_{t+1}\}$ and to a bound ν_τ on the quantity $\ln\big(\max_{t\le\tau}\Pi_{t+1}\big)$. For any single value of $\mathcal{L}'_\tau$, ν_τ is doubled $\mathcal{O}(\ln\nu_T)$ times, and the total regret of runs is $\mathcal{O}\Big(\big(1 + d_T + \mathcal{A}\big)\ln\nu_T + \nu_T + \sqrt{\mathcal{L}'_\tau\nu_T}\Big)$, where we denote $\mathcal{A} = \max_{t\le T}\{\mathcal{A}_{t+1}\}$. Adding up these values for all doubled values of $\mathcal{L}'_\tau$ yields a regret bound of

$$R_T = \mathcal{O}\left(\left(\big(1 + d_T + \mathcal{A}\big)\ln\nu_T + \nu_T\right)\ln\mathcal{L}'_{T+1} + \sqrt{\mathcal{L}'_{T+1}\nu_T}\right) \ . \tag{1}$$

Note that $\mathcal{L}'_{T+1} \le 2\max_{t\le T}\{L_t^* + \mathcal{A}_{t+1}\} \le 2(L_T^* + \mathcal{A})$, where the first inequality holds since $\mathcal{L}'_{T+1}$ is doubled only until it exceeds $\max_{t\le T}\{L_t^* + \mathcal{A}_{t+1}\}$. This gives $\sqrt{(\mathcal{L}'_{T+1}/2)\nu_T} \le \sqrt{L_T^*\nu_T} + \frac{1}{2}(\mathcal{A} + \nu_T)$. Therefore, (1) gives the following.

Corollary 4 *Let $\mathcal{A} = \max_{t\le T}\{\mathcal{A}_{t+1}\}$ and $\Pi = \max_{t\le T}\Pi_{t+1}$. Applying a doubling trick to Hed_C yields the regret bound*

$$R_T = \mathcal{O}\left(\Big(1 + \ln\big(L_T^* + \mathcal{A}\big)\Big)\big((d_T + \mathcal{A})\ln\ln\Pi + \ln\Pi\big) + \sqrt{L_T^*\ln\Pi}\right) \ .$$

Finally, we point out that if $K \ge 2$ is the maximal degree of splits in the tree for $t > 1$, then $\Pi \le N_1(2K-2)^{d_T}$, and the main term $\sqrt{L_T^*\ln\Pi}$ in the above regret bound becomes of order $\sqrt{\big(\ln N_1 + d_T\ln K\big)L_T^*}$.

5. Applications

Few leading experts. We consider a best expert scenario with N experts, where the set of experts that happen to be "leaders" throughout the game is small. The set of all-time leaders (leader set, for short) includes initially only the first expert. In every round the current best expert is added to the set iff its current cumulative loss is strictly smaller than the cumulative loss of all experts *in the leader set*. We generalize this definition by requiring that the advantage over all experts in the leader set must be strictly greater than $\alpha \ge 0$, where α is a parameter. Formally, the leader set starts as $S_1 = \{1\}$, and at the beginning of each round $t > 1$, $S_t = S_{t-1} \cup \{m(t-1)\}$ iff $L_{m(t-1),t-1} + \alpha < L_{j,t-1}$ for every expert

$j \in S_{t-1}$. In adversarial branching terms, we will consider such a new leader as "revealed" at time t. It will branch off the former best expert i in S_{t-1}, namely, the one that satisfies $L_{i,t-2} \leq L_{j,t-2}$ for every $j \in S_{t-1}$ (where the smallest index is taken in a tie). Thus a split will always have two children: a previous leader and the new one.

We denote the number of leaders in the first T steps by $\Lambda_{\alpha,T}$. We assume that upper bounds $\Lambda'_\alpha \geq \Lambda_{\alpha,T}$ and $L' \geq L^*_T$ are known in advance (although the identities of the leaders are not), where a doubling trick is used to guess both quantities. The learner may run Hed_C while simulating an adversary that reveals experts gradually, making Theorem 3 applicable with the following settings: $N_1 = 1$, $d_T = \Lambda_{\alpha,T} - 1$, $\mathcal{A}_{t+1} \leq d_T \max\{\alpha, 1-\alpha\}$, and $\Pi_{t+1} \leq 2^{d_T}$. This provides a bound on the regret w.r.t. the best *revealed* expert. Since the cumulative loss of the best overall expert is smaller by at most $\alpha + 1$, we simply need to add $\alpha + 1$ to this bound to get a bound on the regret. We thus obtain the following.

Theorem 5 *Let $\alpha_1 = \max\{\alpha, 1-\alpha\}$. In the few leading experts scenario, the regret R_T of Hed_C run with parameters $\eta = \ln\Bigl(1 + \sqrt{(2\ln 2)(\Lambda'_\alpha - 1)/(L' + (\Lambda'_\alpha - 1)\alpha_1)}\Bigr)$, $\mathcal{L}' = L' + (\Lambda'_\alpha - 1)\alpha_1$, and $\Pi' = 2^{\Lambda'_\alpha - 1}$ satisfies $R_T = \mathcal{O}\Bigl(\Lambda'_\alpha(\alpha+1) + \sqrt{(\Lambda'_\alpha - 1)L'}\Bigr)$.*

We point out that there is a tradeoff in the choice of α, since an increase in α causes a decrease in $\Lambda_{\alpha,T}$. A doubling trick may again be applied to guess both Λ'_α and L': When either bound is violated, the bound is doubled and the algorithm is restarted.

Corollary 6 *Applying a doubling trick to Hed_C in the few leading experts scenario yields a regret bound of $\mathcal{O}\Bigl(\Lambda_{\alpha,T}(\alpha+1)(1+\ln L^*_T) + \sqrt{L^*_T\,\Lambda_{\alpha,T}}\Bigr)$.*

Clustered experts. We next consider a best expert scenario where experts may be divided into a small number of subsets such that the cumulative losses inside each subset are "similar" at all times. Intuitively, working with one representative of each subset instead of the individual experts is a good approximation for the original problem. An important difference is that the number of representatives may be much smaller than the number of experts, making the regret bound better. Given the approximated regret bound, the maximal "diameter" of the subsets may be added to obtain a regret bound for the original problem.

Formally, let $\alpha \geq 0$ be pre-determined by the learner, and let $\mathcal{N}_{\alpha,T}$ be the number of subsets of experts that are α-*similar* after T steps. Namely, for every $t = 1, \ldots, T$ and every experts i and j in the same subset, $|L_{i,t} - L_{j,t}| \leq \alpha$. As before, we start by assuming we know upper bounds $\mathcal{N}'_\alpha \geq \mathcal{N}_{\alpha,T}$ and $L' \geq L^*_T$, and eventually relax this assumption using a doubling trick.

The learner will implement Hed_C in conjunction with the following splitting scheme. Initially, all experts reside in the same cluster. For every t, a cluster is split at the beginning of time $t+1$ iff the difference between the cumulative losses of any two experts inside it at time t exceeds $\beta = (2\mathcal{N}'_\alpha - 1)\alpha$. To split a cluster, we first sort its members by their cumulative loss at time t. Then we find the largest gap between cumulative loss values and split there. (If more than one maximal gap exists, we pick one arbitrarily.) Since $\beta \geq (2\mathcal{N}_{\alpha,T} - 1)\alpha$ and all the subsets are α-similar, this gap must be larger than α. Furthermore, members of any given subset cannot be on both sides of the gap. Next, if the gap in either of the two parts is larger than β, the process is repeated. Thus, β is an upper

bound on the diameter of each cluster (i.e., the largest difference $|L_{i,T} - L_{j,T}|$ over pairs of experts i, j in the cluster) after any number T of steps. In addition, since clusters always contain entire subsets, the total number of splits after T steps does not exceed $\mathcal{N}_{\alpha,T} - 1$.

We apply Theorem 3 with the settings $N_1 = 1$, $d_T \le \mathcal{N}'_\alpha - 1$, and $\mathcal{A}_{t+1} \le \beta d_T \le (2\mathcal{N}'_\alpha - 1)(\mathcal{N}'_\alpha - 1)\,\alpha$. As for Π', recall that it upper bounds $\Pi_{T+1} = \prod_{\tau=1}^{T} \max\{1, 2|C(i_\tau, \tau+1)| - 2\}$, since $N_1 = 1$. Let $n_1, \ldots, n_k$ be the values of $2|C(i_\tau, \tau+1)| - 2$ in the product that are not zero. We have

$$\Pi_{T+1} = \prod_{i=1}^{k} n_i \le \left(\frac{1}{k}\sum_{i=1}^{k} n_i\right)^k \le \left(\frac{2\mathcal{N}_{\alpha,T} - 2}{k}\right)^k \le \exp\{(2/e)(\mathcal{N}_{\alpha,T} - 1)\}$$

where the last inequality holds since the function $(a/x)^x$ is maximized at $x = a/e$ for every $a > 0$. We may thus set $\Pi' = \exp\{(2/e)(\mathcal{N}'_\alpha - 1)\}$. Theorem 3 now yields a bound on the regret w.r.t. the best revealed expert. We still need to add to this bound the quantity $\beta + 1$, which bounds the difference between the cumulative losses of the best revealed expert and the best overall expert. We obtain the following result for the case of α-similar subsets.

Theorem 7 *In the clustered experts scenario, the regret R_T of* Hed_C *run with parameters* $\eta = \ln\left(1 + \sqrt{(4/e)(\mathcal{N}'_\alpha - 1)/\left(L' + (2\mathcal{N}'_\alpha - 1)(\mathcal{N}'_\alpha - 1)\alpha\right)}\right)$ *and* $\mathcal{L}' = L' + (2\mathcal{N}'_\alpha - 1)(\mathcal{N}'_\alpha - 1)\alpha$ *satisfies* $R_T = \mathcal{O}\left(\mathcal{N}'_\alpha(1 + (2\mathcal{N}'_\alpha - 1)\alpha) + \sqrt{L'(\mathcal{N}'_\alpha - 1)}\right)$.

If both $\mathcal{N}'_\alpha$ and L' are unknown, a doubling trick once again may be used.

Corollary 8 *Applying a doubling trick to* Hed_C *in the clustered experts setting yields a regret bound of* $\mathcal{O}\left(\mathcal{N}_{\alpha,T}(1 + \alpha\mathcal{N}_{\alpha,T})(1 + \ln L^*_T) + \sqrt{L^*_T \mathcal{N}_{\alpha,T}}\right)$.

Remark 9 *If losses contain random noise, the diameter of a set of experts grows gradually over time, rather than remaining constant. Fix a time horizon T and consider N experts with i.i.d. Bernoulli random losses. For $\delta \in (0,1)$, the diameter of this set is $O(\sqrt{T \ln(N/\delta)})$ with probability at least $1 - \delta$. This is shown by combining a "maximal" concentration inequality with the union bound. Picking $\alpha = \Theta(\sqrt{T \ln(N/\delta)})$ for this case thus yields a single cluster and $O(\sqrt{T \ln(N/\delta)})$ regret. A similar argument applies to the few leading experts scenario.*

6. Lower bounds

In this section we prove lower bounds for the branching setup, as well as the few leaders and clustered experts scenarios of Section 5. We show that the key term in the regret bound of Hed_C for the branching setting, $\sqrt{L^*_T \ln \Pi}$ (see Corollary 4), may not be improved in general. The same holds for the corresponding terms $\sqrt{L^*_T \Lambda_{\alpha,T}}$ and $\sqrt{L^*_T \mathcal{N}_{\alpha,T}}$ for the other two scenarios (Corollaries 6 and 8, respectively) if the number of leaders or similar subsets is at most logarithmic in the number of experts. This condition is clearly necessary, since otherwise Hedge itself guarantees better regret than Hed_C.

We use a single construction for all the above scenarios. It involves an oblivious stochastic adversary whose branching tree is a highly unbalanced comb-shaped tree, that is, with splits occurring only in a single branch. This construction and accompanying lemma are geared towards the case of subsets of identical experts, but are useful for the other scenarios

as well. The construction proceeds as follows. Given N for the number of experts and K for the number of unique experts, we define sets $S_1 \supset S_2 \supset \cdots \supset S_K$ of experts, where $S_1 = \{1, \ldots, N\}$, and S_{i+1} is a random half of S_i, which contains the best expert. The K distinct subsets are $S_j \setminus S_{j+1}$, for $j = 1, \ldots, K$, where we define $S_{K+1} = \emptyset$. Just as in proofs for the standard best expert setting (see, e.g., Theorems 4.7 and 4.8 in Chapter 4 of (Nisan et al., 2007)), this construction prevents any learner from doing better than random. We comment, however, that additional care is required to control the number of distinct experts. We make use of the following lemma, proven in the appendix.

Lemma 10 *Let $\ell_{i,t} \in \{0, 1\}$ for all i and t, where exactly K loss sequences $(\ell_{i,1}, \ldots, \ell_{i,T})$ are distinct. Even if K is known to the learner, it holds that*

(i) For every $T \leq \lfloor \log_2 N \rfloor$, there is an oblivious stochastic adversary that generates N loss sequences of length T of which $K = T+1$ are unique, such that the expected regret of any algorithm satisfies $\mathbb{E}[R_T] \geq T/2$.

(ii) For every $T \geq 1 + \lfloor \log_2 N \rfloor$ and $K \leq 1 + \lfloor \log_2 N \rfloor$ there is an oblivious stochastic adversary that generates N loss sequences of length T, of which K are unique, such that the expected regret of any algorithm satisfies $\mathbb{E}[R_T] \geq \frac{1}{4}\sqrt{\lfloor T/(K-1) \rfloor}(K-1) = \Omega\big(\sqrt{T(K-1)}\big)$.

With the above lemma handy, we start by considering the branching setup. Given T and $K < T$, the adversary may generate $N = 2^{T-1}$ loss sequences of length T according to part (ii) of Lemma 10. At time t it maintains sets of experts with identical histories up until time t, according to the stochastic construction. These sets are the leaves of its branching tree, which is a comb-shaped tree with $K-1$ binary splits along a single splitting branch, so we have $\Pi = 2^{K-1}$. The adversary will be extra helpful and reveal N to the learner in advance, and also reveal at each time t the current composition of the sets in the leaves. Even so, by Lemma 10, the regret of any algorithm satisfies $\mathbb{E}[R_T] = \Omega\big(\sqrt{T(K-1)}\big) = \Omega\big(\sqrt{L_T^* \ln \Pi}\big)$, so the key term $\sqrt{L_T^* \ln \Pi}$ in the regret bound for Hed_C (Corollary 4) may not be improved in general.

For α-similar subsets, Lemma 10 clearly gives an $\Omega\big(\sqrt{L_T^*\,(\mathcal{N}_{0,T} - 1)}\big)$ expected regret bound in the $\alpha = 0$ case, if $\mathcal{N}_{0,T} \leq 1 + \lfloor \log_2 N \rfloor \leq T$. Since the adversary is oblivious, we also have $\mathcal{N}_{0,T} \geq \mathcal{N}_{\alpha,T}$ and therefore an $\Omega\big(\sqrt{L_T^*\,(\mathcal{N}_{\alpha,T} - 1)}\big)$ bound.

Finally, we may show that for the few leaders scenario, if $\Lambda_{0,T} \leq 1 + \lfloor \log_2 N \rfloor \leq cT$, for some $c > 0$, then the expected regret is $\Omega\big(\sqrt{L_T^*\,(\Lambda_{0,T} - 1)}\big)$. The application of Lemma 10 for this case requires an additional technical step. A closer examination of our stochastic construction reveals that if a run with $K \leq \log_2 N$ is stopped at time $T/2$, then the expected regret is still $\Omega\big(\sqrt{L_T^*\,(K-1)}\big)$, the number of leaders is in $\{1, \ldots, K\}$, and the set of best experts is of size $\Theta\big(\sqrt{N}\big)$. We may then artificially raise the number of leaders to $K \leq \log_2 N = O\big(\sqrt{N}\big)$ by sequentially giving ϵ loss to one member of the best expert set, and a round later to all the others. The remaining rounds may be filled with zero losses for all experts. Thus, for any $\Lambda \leq \log_2 N$, if we run this modified procedure with $K = \Lambda_{0,T}$, we achieve expected regret of $\Omega\big(\sqrt{L_T^*\,(\Lambda_{0,T} - 1)}\big)$ and the number of leaders is exactly $\Lambda_{0,T}$. As before, this implies an $\Omega\big(\sqrt{L_T^*\,(\Lambda_{\alpha,T} - 1)}\big)$ as well. The following theorem summarizes our lower bounds.

Theorem 11 *Let T be a known time horizon.*

(i) *For the branching setting and for any $d < T$, there is a random tree generation with $d_T = d$ such that the expected regret of any algorithm satisfies $\mathbb{E}[R_T] = \Omega\big(\sqrt{T \ln \Pi}\big)$.*

(ii) *For any $\mathcal{N}_{0,T} \le 1 + \lfloor \log_2 N \rfloor \le T$, there is a random construction of N experts of which $\mathcal{N}_{0,T}$ are distinct, such that the expected regret of any algorithm satisfies $\mathbb{E}[R_T] = \Omega\big(\sqrt{T(\mathcal{N}_{0,T} - 1)}\big)$.*

(iii) *For any $\Lambda_{0,T} \le 1 + \lfloor \log_2 N \rfloor \le cT$, for some $c > 0$, there is a random construction of N experts of which $\Lambda_{0,T}$ are leaders, such that the expected regret of any algorithm satisfies $\mathbb{E}[R_T] = \Omega\big(\sqrt{T(\Lambda_{0,T} - 1)}\big)$.*

7. Branching experts for the multi-armed bandit setting

In this section we introduce and analyze a variant of the randomized multi-armed bandit algorithm Exp3 of Auer et al. (2002) for the branching setting. For the sake of simplicity, we focus on the case of perfect cloning. This means that new actions $j \in C(i, t+1)$ all start off with the same cumulative loss $L_{i,t}$ as their parent i. This variant, called Exp3.C, is described below here.

Branching Exp3 (Exp3.C)

Parameters: A sequence $\eta_1, \eta_2, \ldots$ of real-valued functions satisfying the assumptions of Theorem 12.

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

1. For each action $i = 1, \ldots, N_{t-1}$, after the adversary reveals the set $C(i,t)$:
   If $t = 1$, then let $\widetilde{L}_{j,0} = 0$ for every $j = 1, \ldots, N_1$;
   else, if $t > 1$, then $\widetilde{L}_{j,t-1} = \widetilde{L}_{i,t-2} + \widetilde{\ell}_{i,t-1} + \frac{1}{\eta_{t-1}} \ln |C(i,t)|$ for every $j \in C(i,t)$, including i.
2. Compute the new distribution over actions $\boldsymbol{p}_t = \big(p_{1,t}, \ldots, p_{N_t,t}\big)$, where
$$p_{i,t} = \frac{\exp\left(-\eta_t \widetilde{L}_{i,t-1}\right)}{\sum_{k=1}^{N_t} \exp\left(-\eta_t \widetilde{L}_{k,t-1}\right)} \,.$$
3. Draw an action I_t from the probability distribution $\boldsymbol{p}_t$ and observe loss $\ell_{I_t,t}$.
4. For each action $i = 1, \ldots, N_t$ compute the estimated loss $\widetilde{\ell}_{i,t} = \dfrac{\ell_{i,t}}{p_{i,t}} \mathbb{I}\{I_t = i\}$.

The main modification with respect to Exp3 is in the way cumulative loss estimates $\widetilde{L}_{i,t}$ are computed (step 1 in the pseudo-code). The additional term $\frac{1}{\eta_{t-1}} \ln |C(i,t)|$ in these estimates serves a role similar to that of the partial restart in the full information case. There, we divided the weight of a parent expert i among children $C(i,t)$. Here, we increase the loss estimate of $j \in C(i,t)$ to achieve the same effect.

The next theorem bounds the expected regret $\mathbb{E}\, R_T$ of Exp3.C against an oblivious adversary.[5] This is defined as $\mathbb{E}\, R_T = \mathbb{E}\big[\widehat{L}_T\big] - L_T^*$, where $\widehat{L}_T = \ell_{I_1,1} + \cdots + \ell_{I_T,T}$ is the

5. Extensions to nonoblivious adversaries are possible, with some assumptions on the adversary's control on the quantities $C(i,t)$.

random variable denoting Exp3.C's total loss with respect to the sequence $I_1, \ldots, I_T$ of random draws.

Theorem 12 *Let $\eta_1, \eta_2, \ldots$ be a sequence of functions $\eta_t : \mathbb{N} \to \mathbb{R}^+$ such that for every $k_1 \leq k_2 \leq \ldots$, it holds that $\eta_1(k_1) \geq \eta_2(k_2) \geq \ldots$ (in what follows, we write $\eta_t = \eta_t(N_t)$ for short). If $i_0, \ldots, i_T = m(T)$ are the actions on the path from the root to the best action $m(T)$, then*

$$\mathbb{E}\, R_T \leq \frac{1}{2} \sum_{t=1}^{T} N_t \eta_t + \sum_{t=1}^{T} \frac{1}{\eta_t} \ln \frac{N_t |C(i_t, t+1)|}{N_{t+1}} + \frac{\ln N_{T+1}}{\eta_T} \ . \tag{2}$$

If Exp3.C is run with $\eta_t(k) = \sqrt{\frac{\ln ek}{tk}}$, then

$$\mathbb{E}\, R_T \leq 2\sqrt{T N_T \ln e N_T} \left(1 + \frac{\ln \prod_{t=1}^{T} |C(i_t, t+1)|}{2 \ln e N_T} \right) \ . \tag{3}$$

Acknowledgments

We wish to thank the anonymous reviewers for their helpful comments. This research was supported in part by the Google Inter-university center for Electronic Markets and Auctions, by a grant from the Israel Science Foundation, by a grant from United States-Israel Binational Science Foundation (BSF), by a grant from the Israeli Ministry of Science (MoS), and by The Israeli Centers of Research Excellence (I-CORE) program (Center No. 4/11). This work is part of Ph.D. thesis research carried out by the first author at Tel Aviv University.

References

Peter Auer, Nicolò Cesa-Bianchi, Yoav Freund, and Robert E. Schapire. The nonstochastic multiarmed bandit problem. *SIAM Journal on Computing*, 32(1):48–77, 2002.

Avrim Blum and Yishay Mansour. From external to internal regret. *Journal of Machine Learning Research*, 8:1307–1324, 2007.

Olivier Bousquet and Manfred Warmuth. Tracking a small set of experts by mixing past posteriors. *Journal of Machine Learning Research*, 3:363–396, 2002.

Sébastien Bubeck and Nicolò Cesa-Bianchi. Regret analysis of stochastic and nonstochastic multi-armed bandit problems. *Foundations and Trends in Machine Learning*, 5(1):1–122, 2012.

N. Cesa-Bianchi, Y. Mansour, and G. Stoltz. Improved second-order bounds for prediction with expert advice. *Machine Learning*, 66(2-3):321–352, 2007.

Nicolò Cesa-Bianchi and Gábor Lugosi. *Prediction, Learning, and Games*. Cambridge University Press, 2006.

Kamalika Chaudhuri, Yoav Freund, and Daniel Hsu. A parameter-free hedging algorithm. In *NIPS*, pages 297–305, 2009.

Alexey V. Chernov and Vladimir Vovk. Prediction with advice of unknown number of experts. In *UAI*, pages 117–125, 2010.

Chao-Kai Chiang, Tianbao Yang, Chia-Jung Lee, Mehrdad Mahdavi, Chi-Jen Lu, Rong Jin, and Shenghuo Zhu. Online optimization with gradual variations. *Journal of Machine Learning Research - Proceedings Track*, 23:6.1–6.20, 2012.

Peter DeMarzo, Ilan Kremer, and Yishay Mansour. Online trading algorithms and robust option pricing. In *STOC*, pages 477–486, 2006.

Y. Freund, R. Schapire, Y. Singer, and M. Warmuth. Using and combining predictors that specialize. In *Proceedings of the 29th Annual ACM Symposium on the Theory of Computing*, pages 334–343. ACM Press, 1997.

Yoav Freund and Robert E. Schapire. A decision-theoretic generalization of on-line learning and an application to boosting. In *Euro-COLT*, pages 23–37. Springer-Verlag, 1995.

E. Gofer and Y. Mansour. Pricing exotic derivatives using regret minimization. In *Proc. of the 4th Symposium on Algorithmic Game Theory (SAGT)*, 2011.

E. Hazan and S. Kale. Extracting certainty from uncertainty: regret bounded by variation in costs. In *COLT*, pages 57–68, 2008.

M. Herbster and M.K. Warmuth. Tracking the best expert. *Machine Learning*, 32(2): 151–178, 1998.

A. Kalai and S. Vempala. Efficient algorithms for online decision problems. *Journal of Computer and System Sciences*, 71(3):291–307, 2005.

R. Kleinberg, A. Niculescu-Mizil, and Y. Sharma. Regret bounds for sleeping experts and bandits. *Machine learning*, 80(2):245–272, 2010.

Wouter Koolen, Dmitry Adamskiy, and Manfred Warmuth. Putting Bayes to sleep. In *NIPS*, pages 135–143, 2012.

N. Littlestone and M.K. Warmuth. The weighted majority algorithm. *Information and Computation*, 108:212–261, 1994.

Noam Nisan, Tim Roughgarden, Eva Tardos, and Vijay V. Vazirani. *Algorithmic Game Theory*. Cambridge University Press, New York, NY, USA, 2007. ISBN 0521872820.

Alexander Rakhlin, Ohad Shamir, and Karthik Sridharan. Relax and localize: From value to algorithms. *arXiv preprint arXiv:1204.0870*, 2012.

Cosma Rohilla Shalizi, Abigail Z. Jacobs, and Aaron Clauset. Adapting to non-stationarity with growing expert ensembles. *CoRR*, abs/1103.0949, 2011.

V.G. Vovk. Aggregating strategies. In *Proceedings of the 3rd Annual Workshop on Computational Learning Theory*, pages 372–383, 1990.

Appendix A. Missing proofs

Proof of Lemma 1: Recall that $w_{i,t+1} = w'_{i,t}e^{-\eta\ell_{i,t}}$ and $W'_t = W_t$ for every $i = 1,\ldots,N_{t+1}$, $t = 1,\ldots,T$ and note that $W'_t > 0$. Let $p'_{i,t} = w'_{i,t}/W'_t$ for all $i = 1,\ldots,N_{t+1}$, so that $\boldsymbol{p}'_t = (p'_{1,t},\ldots,p'_{N_{t+1},t})$ is a probability vector. Hence, for every t,

$$\ln\frac{W_{t+1}}{W_t} = \ln\frac{\sum_{i=1}^{N_{t+1}} w_{i,t+1}}{W'_t} = \ln\frac{\sum_{i=1}^{N_{t+1}} w'_{i,t}e^{-\eta\ell_{i,t}}}{W'_t} = \ln\sum_{i=1}^{N_{t+1}} p'_{i,t}e^{-\eta\ell_{i,t}} \ .$$

Now, if there is no restart at time $t+1$, that is $|C(i,t+1)| = 1$ for all $i = 1,\ldots,N_t$, then $N_{t+1} = N_t$, $w'_{i,t} = w_{i,t}$, and $p'_{i,t} = p_{i,t}$, $i = 1,\ldots,N_t$. In this case, by the convexity of $f(x) = e^x$, we have $e^{-\eta\ell_{i,t}} \le 1 - \big(1-e^{-\eta}\big)\,\ell_{i,t}$. Hence we can write

$$\begin{aligned}
\ln\sum_{i=1}^{N_{t+1}} p'_{i,t}e^{-\eta\ell_{i,t}} &= \ln\sum_{i=1}^{N_t} p_{i,t}e^{-\eta\ell_{i,t}} \\
&\le \ln\sum_{i=1}^{N_t} p_{i,t}\big(1-(1-e^{-\eta})\ell_{i,t}\big) \\
&\le -(1-e^{-\eta})\boldsymbol{p}_t\cdot\boldsymbol{\ell}_t \\
&= -(1-e^{-\eta})\widehat{\ell}_t \ .
\end{aligned}$$

On the other hand, if a restart takes place, then we have

$$\ln\sum_{i=1}^{N_{t+1}} p'_{i,t}e^{-\eta\ell_{i,t}} \le \ln\sum_{i=1}^{N_{t+1}} p'_{i,t}\big(1-(1-e^{-\eta})\ell_{i,t}\big) \le -(1-e^{-\eta})\sum_{i=1}^{N_{t+1}} p'_{i,t}\ell_{i,t}$$

which is trivially upper bounded by $-(1-e^{-\eta})(\widehat{\ell}_t - 1)$, since $\sum_{i=1}^{N_{t+1}} p'_{i,t}\ell_{i,t} \ge 0 \ge \widehat{\ell}_t - 1$.

Thus, if $n' \le n$ is the number of restarts,

$$\ln\frac{W_{T+1}}{W_1} = \sum_{t=1}^{T}\ln\frac{W_{t+1}}{W_t} \le -(1-e^{-\eta})\big(\widehat{L}_T - n'\big) \le -(1-e^{-\eta})\big(\widehat{L}_T - n\big) \ .$$

■

Proof of Lemma 2: Recall that $i_0,\ldots,i_t$ is a path from the root i_0 to expert i_t in the branching experts tree, where $i_\tau \in C(i_{\tau-1},\tau)$ for all $\tau = 1,\ldots,t$. We first prove by induction that

$$w_{i_t,t} \ge e^{-\eta L_{i_t,t-1}}\prod_{\tau=1}^{t-1}\Big(1+\big(|C(i_\tau,\tau+1)|-1\big)e^{\eta\alpha_{i_\tau,\tau}}\Big)^{-1} \ .$$

For $t = 1$ both sides equal 1 and the claim is trivial. We next assume the claim holds for t and prove it for $t+1$. For every t we have that

$$\begin{aligned}
\sum_{j\in C(i_t,t+1)} e^{-\eta L_{j,t-1}} &= e^{-\eta L_{i_t,t-1}}\sum_{j\in C(i_t,t+1)} e^{-\eta(L_{j,t-1}-L_{i_t,t-1})} \\
&\le e^{-\eta L_{i_t,t-1}}\Big(1+\big(|C(i_t,t+1)|-1\big)e^{\eta\alpha_{i_t,t}}\Big)
\end{aligned}$$

where we used $\alpha_{i_t,t} = \max\left\{\left|L_{j,t-1} - L_{k,t-1}\right| : j,k \in C(i_t,t+1)\right\}$ and also $i_t \in C(i_t,t+1)$. Thus

$$\begin{aligned} w_{i_{t+1},t+1} &= w'_{i_{t+1},t} e^{-\eta \ell_{i_{t+1},t}} = w_{i_t,t} \frac{e^{-\eta L_{i_{t+1},t}}}{\sum_{j \in C(i_t,t+1)} e^{-\eta L_{j,t-1}}} \\ &\geq w_{i_t,t} e^{-\eta\left(L_{i_{t+1},t} - L_{i_t,t-1}\right)} \left(1 + \left(|C(i_t,t+1)| - 1\right) e^{\eta \alpha_{i_t,t}}\right)^{-1} \\ &\geq e^{-\eta L_{i_{t+1},t}} \prod_{\tau=1}^{t} \left(1 + \left(|C(i_\tau,\tau+1)| - 1\right) e^{\eta \alpha_{i_\tau,\tau}}\right)^{-1} \end{aligned}$$

completing the induction. It is easy to verify that

$$1 + \left(|C(i_\tau,\tau+1)| - 1\right) e^{\eta \alpha_{i_\tau,\tau}} \leq \max\left\{1, 2|C(i_\tau,\tau+1)| - 2\right\} e^{\eta \alpha_{i_\tau,\tau}}$$

and thus

$$\begin{aligned} w_{i_t,t} &\geq e^{-\eta L_{i_t,t-1}} \prod_{\tau=1}^{t-1} \left(\max\left\{1, 2|C(i_\tau,\tau+1)| - 2\right\} e^{\eta \alpha_{i_\tau,\tau}}\right)^{-1} \\ &= \exp\left(-\eta L_{i_t,t-1} - \eta \sum_{\tau=1}^{t-1} \alpha_{i_\tau,\tau}\right) \prod_{\tau=1}^{t-1} \max\left\{1, 2|C(i_\tau,\tau+1)| - 2\right\}^{-1} . \end{aligned}$$

■

Proof of Theorem 3: By Lemma 2 we have that

$$\ln \frac{W_{T+1}}{W_1} \geq \ln \frac{w_{m(T),T+1}}{N_1} \geq -\eta L^*_T - \eta \mathcal{A}_{T+1} - \ln \Pi_{T+1} \geq -\eta\left(L^*_T + \mathcal{A}_{T+1}\right) - \ln \Pi' .$$

In addition, by Lemma 1, $\ln(W_{T+1}/W_1) \leq -(1 - e^{-\eta})(\widehat{L}_T - d_T)$. Combining the upper and lower bounds for $\ln(W_{T+1}/W_1)$ and rearranging, we have

$$\widehat{L}_T - d_T \leq \frac{\eta\left(L^*_T + \mathcal{A}_{T+1}\right) + \ln \Pi'}{1 - e^{-\eta}} .$$

The rest of the proof is similar to the standard proof of Hedge and is given here for completeness. Denote $\nu = \ln \Pi'$ and $\mathcal{L} = L^*_T + \mathcal{A}_{T+1}$. It is easily verified that $\eta \leq \frac{1}{2}(e^\eta - e^{-\eta})$ for every $\eta \geq 0$, and therefore,

$$\widehat{L}_T - d_T \leq \frac{\frac{1}{2}(e^\eta - e^{-\eta})\mathcal{L} + \nu}{1 - e^{-\eta}} = \frac{\nu}{1 - e^{-\eta}} + \frac{\mathcal{L}}{2}(e^\eta + 1) .$$

Recall that we set η such that $e^\eta = 1 + \sqrt{2\nu/\mathcal{L}'}$, and therefore

$$\begin{aligned} R_T &= \widehat{L}_T + \mathcal{A}_{T+1} - \mathcal{L} \leq d_T + \mathcal{A}_{T+1} + \frac{\nu}{1 - e^{-\eta}} + \frac{\mathcal{L}}{2}(e^\eta - 1) \\ &\leq d_T + \mathcal{A}_{T+1} + \frac{\nu}{1 - e^{-\eta}} + \frac{\mathcal{L}'}{2}(e^\eta - 1) \\ &= d_T + \mathcal{A}_{T+1} + \left(1 + \sqrt{\mathcal{L}'/(2\nu)}\right)\nu + \frac{\mathcal{L}'}{2}\sqrt{2\nu/\mathcal{L}'} \\ &= d_T + \mathcal{A}_{T+1} + \nu + \sqrt{2\mathcal{L}'\nu} \end{aligned}$$

completing the proof. ■

Proof of Lemma 10: (i) Consider the following loss sequence generation. At each time t, the adversary focuses only on the set S_t of experts, where $S_t = \{1 \le i \le 2^{\lfloor \log_2 N \rfloor} : \forall \tau < t, \ell_{i,\tau} = 0\}$ (initially, S_1 includes all the experts $1 \le i \le 2^{\lfloor \log_2 N \rfloor}$). The set S_t is then divided randomly into two equal-sized sets, where the first set is given loss 1 at time t and the other loss 0. All other experts are given loss 1. Even if the learner has the benefit of knowing S_t before deciding on $\boldsymbol{p}_t$, its expected loss at time t is 1/2. (The learner only stands to lose if it puts weight outside S_t.) Therefore, in any case, $\mathbb{E}[\widehat{L}_T] \ge T/2$, while $L^*_T = 0$, implying that $\mathbb{E}[R_T] \ge T/2$. There are exactly $T+1$ distinct experts in the above construction.

(ii) Consider the following construction of loss sequences by an adversary. Let $\tau = \lfloor T/(K-1) \rfloor$ and divide the time range $1, \ldots, \tau(K-1)$ into $K-1$ time slices of size τ. Times $(\tau(K-1), T]$ may be ignored, since the adversary may assign $\ell_{i,t} = 0$ for every $1 \le i \le N$ and $t \in (\tau(K-1), T]$, so the regret of any algorithm is unaffected. The adversary defines sets $S_1 \supset S_2 \supset \cdots \supset S_K$ of experts, where initially $S_1 = \{1, \ldots, N\}$. For time slice j, all experts not in S_j incur τ times the loss 1. Denote S^1_j for the first $\lfloor |S_j|/2 \rfloor$ experts in S_j and $S^2_j = S_j \setminus S^1_j$. The adversary generates two $\{0,1\}$-valued loss sequences of size τ by making 2τ i.i.d. draws from the uniform distribution on $\{0,1\}$. Experts in S^1_j incur the losses in the first sequence, and experts in S^2_j incur the losses in the second sequence. If the sequences are identical, the adversary modifies its choice by picking instead the sequences $\{0, \ldots, 0, 1\}$ and $\{0, \ldots, 0, 0\}$ and assigning one of them randomly to the experts in S^1_j and the other to experts in S^2_j. S_{j+1} is then defined as the set between S^1_j and S^2_j with the smallest cumulative loss, or S^1_j in case of a tie. Note that we end up with exactly K distinct loss sequences, and S_j always contains an expert with the smallest cumulative loss at any time.

We may assume w.l.o.g. that in time slice j, the algorithm puts weight only on S_j, and we denote R^j for its regret on the j-th time slice. We also denote $\widehat{R}^j$ for the regret if we had not made the modification for identical sequences. By Lemma 14 (see below), we have $\mathbb{E}[\widehat{R}^j] \ge \sqrt{\tau}/4$, since the expected loss of the algorithm on the time slice is $\tau/2$. Denote B_j for the event that the sequences in slice j are identical and $\bar{B}_j$ for its complement. We have $\sqrt{\tau}/4 \le \mathbb{E}[\widehat{R}^j] = \mathbb{E}[\widehat{R}^j \mid \bar{B}_j]\mathbb{P}(\bar{B}_j) + \mathbb{E}[\widehat{R}^j \mid B_j]\mathbb{P}(B_j) = \mathbb{E}[\widehat{R}^j \mid \bar{B}_j]\mathbb{P}(\bar{B}_j)$, since the regret is 0 if the sequences are identical. In addition, we have

$$\begin{aligned}
\mathbb{E}[R^j] &= \mathbb{E}[R^j \mid \bar{B}_j]\mathbb{P}(\bar{B}_j) + \mathbb{E}[R^j \mid B_j]\mathbb{P}(B_j) \\
&= \mathbb{E}[\widehat{R}^j \mid \bar{B}_j]\mathbb{P}(\bar{B}_j) + \mathbb{E}[R^j \mid B_j]\mathbb{P}(B_j) \\
&\ge \mathbb{E}[\widehat{R}^j \mid \bar{B}_j]\mathbb{P}(\bar{B}_j) \ge \sqrt{\tau}/4
\end{aligned}$$

where the first inequality is true because when B_j occurs, $R^j \ge 0$ (one sequence is all zeros). S_j always contains an expert with minimal cumulative loss, so

$$\mathbb{E}[R_T] = \sum_{j=1}^{K-1} \mathbb{E}[R^j] \ge \frac{1}{4}\sqrt{\tau}(K-1) = \frac{1}{4}\sqrt{\lfloor T/(K-1) \rfloor}(K-1) = \Omega\big(\sqrt{T(K-1)}\big) \ .$$

■

Proof of Theorem 12: Note first that if splits always occur uniformly for all actions i (i.e, for

all t we have that $|C(i,t+1)|$ is the same for all $i=1,\dots,N_t$), then $N_{t+1}=N_t|C(i,t+1)|$ implying $\ln\big(N_t|C(i,t+1)|/N_{t+1}\big)=0$. Hence we would get

$$\mathbb{E}\,R_T \le \frac{1}{2}\sum_{t=1}^{T} N_t\eta_t + \frac{\ln N_{T+1}}{\eta_T}\ .$$

In particular, for the standard bandit setting, where $N_1=\cdots=N_{T+1}=N$, we get

$$\mathbb{E}\,R_T \le \frac{N}{2}\sum_{t=1}^{T}\eta_t + \frac{\ln N}{\eta_T}$$

and recover the original result.

To prove (3) from (2), we first note that $\ln\frac{N_t}{N_{t+1}}\le 0$ for all t. In addition, without loss of generality, $N_T=N_{T+1}$ (otherwise we add an artificial round). We obtain

$$\begin{aligned}
\mathbb{E}\,R_T &\le \frac{1}{2}\sum_{t=1}^{T}\sqrt{\frac{N_t\ln eN_t}{t}} + \frac{1}{\eta_T}\sum_{t=1}^{T}\ln|C(i_t,t+1)| + \sqrt{\frac{TN_T(\ln N_T)^2}{\ln eN_T}} \\
&\le \frac{1}{2}\sqrt{N_T\ln eN_T}\sum_{t=1}^{T}\frac{1}{\sqrt{t}} + \sqrt{\frac{TN_T}{\ln eN_T}}\ln\prod_{t=1}^{T}|C(i_t,t+1)| + \sqrt{TN_T\ln eN_T} \\
&\le 2\sqrt{TN_T\ln eN_T} + \sqrt{\frac{TN_T}{\ln eN_T}}\ln\prod_{t=1}^{T}|C(i_t,t+1)| \\
&= 2\sqrt{TN_T\ln eN_T}\left(1+\frac{\ln\prod_{t=1}^{T}|C(i_t,t+1)|}{2\ln eN_T}\right)
\end{aligned}$$

where we used the fact that $\sum_{t=1}^{T}\frac{1}{\sqrt{t}}\le\int_0^T\frac{1}{\sqrt{t}}dt=2\sqrt{T}$.

The proof of (2) is an adaptation of the proof of (Bubeck and Cesa-Bianchi, 2012, Theorem 3.1) —indicated with [BS3.1] for short, which is divided into five steps. Here we just focus on the main differences. In the following, we write $\mathbb{E}_{i\sim p_t}$ to denote the expectation w.r.t. the random draw of i from the distribution p_t specified by the probability vector $\boldsymbol{p}_t=\big(p_{1,t},\dots,p_{N_t,t}\big)$. Moreover, given any action $k\in N_T$, we use k also to index any action i on the path from the root to k. This is OK because, since we have perfect cloning, we have that $L_{k,T}=\ell_{i_1,1}+\cdots+\ell_{i_T,T}$ where $i_1,\dots,i_T$ are the actions on the path from the root i_0 to $i_T=k$.

The first two steps of the proof are identical to [BS3.1]:

$$\sum_{t=1}^{T}\ell_{I_t,t}-\sum_{t=1}^{T}\ell_{k,t}=\sum_{t=1}^{T}\mathbb{E}_{i\sim p_t}\widetilde{\ell}_{i,t}-\sum_{t=1}^{T}\mathbb{E}_{I_t\sim p_t}\widetilde{\ell}_{k,t}\ . \tag{4}$$

Now we rewrite $\mathbb{E}_{i\sim p_t}\widetilde{\ell}_{i,t}$ as follows

$$\mathbb{E}_{i\sim p_t}\widetilde{\ell}_{i,t}=\frac{1}{\eta_t}\ln\mathbb{E}_{i\sim p_t}\exp\Big(-\eta_t\big(\widetilde{\ell}_{i,t}-\mathbb{E}_{k\sim p_t}\widetilde{\ell}_{k,t}\big)\Big)-\frac{1}{\eta_t}\ln\mathbb{E}_{i\sim p_t}\exp\Big(-\eta_t\widetilde{\ell}_{i,t}\Big)\ . \tag{5}$$

Following the second step in the proof of [BS3.1] we obtain

$$\ln \mathbb{E}_{i\sim p_t} \exp\left(-\eta_t(\widetilde{\ell}_{i,t} - \mathbb{E}_{k\sim p_t}\widetilde{\ell}_{k,t})\right) \le \frac{\eta_t^2}{2p_{I_t,t}} \tag{6}$$

Next, we study the second term in (5). This relies on the specific properties of Exp3.C. Let $\Phi_0(\eta) = 0$ and $\Phi_t(\eta) = \frac{1}{\eta}\ln\frac{1}{N_{t+1}}\sum_{i=1}^{N_{t+1}}\exp\left(-\eta\widetilde{L}_{i,t}\right)$. By definition of p_t, and recalling that $\widetilde{L}_{j,t} = \widetilde{L}_{i,t}$ for every $j \in C(i,t+1)$ and every $i = 1,\dots,N_t$, we have

$$\begin{aligned}
-\frac{1}{\eta_t}\ln \mathbb{E}_{i\sim p_t}\exp\left(-\eta_t\widetilde{\ell}_{i,t}\right) &= -\frac{1}{\eta_t}\ln\frac{\sum_{i=1}^{N_t}\exp\left(-\eta_t\widetilde{L}_{i,t-1}\right)\exp\left(-\eta_t\widetilde{\ell}_{i,t}\right)}{\sum_{i=1}^{N_t}\exp\left(-\eta_t\widetilde{L}_{i,t-1}\right)} \\
&= -\frac{1}{\eta_t}\ln\frac{\sum_{i=1}^{N_t}|C(i,t+1)|\exp\left(-\eta_t\widetilde{L}_{i,t}\right)}{\sum_{i=1}^{N_t}\exp\left(-\eta_t\widetilde{L}_{i,t-1}\right)} \\
&= -\frac{1}{\eta_t}\ln\frac{\sum_{i=1}^{N_{t+1}}\exp\left(-\eta_t\widetilde{L}_{i,t}\right)}{\sum_{i=1}^{N_t}\exp\left(-\eta_t\widetilde{L}_{i,t-1}\right)} \\
&= \Phi_{t-1}(\eta_t) - \Phi_t(\eta_t) + \frac{1}{\eta_t}\ln\frac{N_t}{N_{t+1}}\ .
\end{aligned} \tag{7}$$

Putting together (4), (5), (6) and (7) we obtain

$$\sum_{t=1}^{T}\ell_{I_t,t} - \sum_{t=1}^{T}\ell_{k,t} \le \sum_{t=1}^{T}\frac{\eta_t}{2p_{I_t,t}} + \sum_{t=1}^{T}\Big(\Phi_{t-1}(\eta_t) - \Phi_t(\eta_t)\Big) + \sum_{t=1}^{T}\frac{1}{\eta_t}\ln\frac{N_t}{N_{t+1}} - \sum_{t=1}^{T}\mathbb{E}_{I_t\sim p_t}\widetilde{\ell}_{k,t}\ .$$

The first term is easy to bound in expectation since by the rule of conditional expectations we have

$$\mathbb{E}\sum_{t=1}^{T}\frac{\eta_t}{2p_{I_t,t}} = \mathbb{E}\sum_{t=1}^{T}\mathbb{E}_{I_t\sim p_t}\frac{\eta_t}{2p_{I_t,t}} = \frac{1}{2}\sum_{t=1}^{T}N_t\eta_t\ .$$

For the second term we again proceed similarly to the proof of [BS3.1],

$$\sum_{t=1}^{T}\big(\Phi_{t-1}(\eta_t) - \Phi_t(\eta_t)\big) = \sum_{t=1}^{T-1}\big(\Phi_t(\eta_{t+1}) - \Phi_t(\eta_t)\big) - \Phi_T(\eta_T)$$

since $\Phi_0(\eta_1) = 0$. Note that

$$\begin{aligned}
-\Phi_T(\eta_T) &= \frac{\ln N_{T+1}}{\eta_T} - \frac{1}{\eta_T}\ln\left(\sum_{i=1}^{N_{T+1}}\exp\left(-\eta_T\widetilde{L}_{i,T}\right)\right) \\
&\le \frac{\ln N_{T+1}}{\eta_T} - \frac{1}{\eta_T}\ln\left(\exp\left(-\eta_T\widetilde{L}_{k,T}\right)\right) \\
&= \frac{\ln N_{T+1}}{\eta_T} + \sum_{t=1}^{T}\left(\tilde{\ell}_{k,t} + \frac{1}{\eta_t}\ln|C(k,t+1)|\right)
\end{aligned}$$

and thus we have

$$\mathbb{E}\left[\sum_{t=1}^{T}\ell_{I_t,t}-\sum_{t=1}^{T}\ell_{k,t}\right]\leq\frac{1}{2}\sum_{t=1}^{T}N_t\eta_t+\mathbb{E}\sum_{t=1}^{T-1}\big(\Phi_t(\eta_{t+1})-\Phi_t(\eta_t)\big)$$
$$+\sum_{t=1}^{T}\frac{1}{\eta_t}\ln\frac{N_t|C(k,t+1)|}{N_{t+1}}+\frac{\ln N_{T+1}}{\eta_T}\ .$$

The proof is concluded by showing that $\Phi_t'(\eta)\geq 0$. Since, $\eta_{t+1}\leq\eta_t$ this would give $\Phi_t(\eta_{t+1})-\Phi_t(\eta_t)\leq 0$. In fact, the proof of this claim goes along the same lines as [BS3.1], and is therefore omitted. ■

Appendix B. Additional claims

Claim 13 *There exists a scenario with $N+1$ experts s.t. for a large enough N, Hed_C has regret $R_T=\Omega\big(\sqrt{L_T^*N}\big)$ while Hedge has regret $R_T=\mathcal{O}\big(\ln N+\sqrt{L_T^*\ln N}\big)$.*

Proof Consider the following special scenario. The adversary initially reveals $N+1$ experts, one in every round, without any losses, in a comb-shaped branching tree. The adversary next gives losses 0 to the last revealed expert and 1 to the others for t rounds, and afterwards gives loss 1 to all experts for τ rounds. Suppose further that N, t, and τ are known in advance.

We now consider the regret of Hed_C. We have $L_T^*=\tau$, so $\eta=\ln\big(1+\sqrt{(2\ln 2)N/\tau}\big)$, and it is easily seen that the regret is lower bounded by $\big(1-p_{N+1,t+N}\big)t$, where

$$p_{N+1,t+N}=\frac{2^{-N}}{2^{-N}+(1-2^{-N})e^{-\eta t}}=\frac{1}{1+(2^N-1)e^{-\eta t}}\leq\frac{e^{\eta t}}{2^N}=\exp(\eta t-N\ln 2)\ .$$

Taking $t=\big\lfloor(N\ln 2-1)/\eta\big\rfloor$ ensures the regret is lower bounded by $t/2$. Since

$$\frac{N\ln 2-1}{\eta}\geq\frac{N\ln 2-1}{\sqrt{(2\ln 2)N/\tau}}\geq\frac{(N/2)\ln 2}{\sqrt{(2\ln 2)N/\tau}}\geq\sqrt{N\tau(\ln 2)/8}$$

the result follows. ■

Lemma 14 *If $X=Z_1+\ldots+Z_n$ and $Y=Z_{n+1}+\ldots+Z_{2n}$, where Z_i are independent Bernoulli variables with $p=\frac{1}{2}$, then $\mathbb{E}[\min(X,Y)]\leq\frac{1}{2}n-\frac{1}{4}\sqrt{n}$.*

Proof We have that

$$\mathbb{E}[\min(X,Y)]=\mathbb{E}\left[\frac{1}{2}(X+Y-|X-Y|)\right]$$
$$=\frac{1}{2}\mathbb{E}[X]+\frac{1}{2}\mathbb{E}[Y]-\frac{1}{2}\mathbb{E}[|X-Y|]$$
$$=\frac{n}{2}-\frac{1}{2}\mathbb{E}[|X-Y|]\ ,$$

so we need only determine the value of $\mathbb{E}[|X - Y|]$. Now, $\sigma_i = 2Z_i - 1$, for $1 \leq i \leq 2n$, are independent Rademacher variables, and $\frac{1}{2}(\sigma_j - \sigma_{j+n}) = Z_j - Z_{j+n}$, for $1 \leq j \leq n$. Thus, $|X - Y| = |\sum_{j=1}^{n}(Z_j - Z_{j+n})| = \frac{1}{2}|\sum_{i=1}^{2n} a_i\sigma_i|$, where $|a_i| = 1$ for every i. By Khinchine's inequality (see, e.g., Cesa-Bianchi and Lugosi (2006), Lemma A.9),

$$\mathbb{E}\left[\left|\sum_{i=1}^{2n} a_i\sigma_i\right|\right] \geq \frac{1}{\sqrt{2}}\sqrt{\sum_{i=1}^{2n} a_i^2} = \sqrt{n} \ .$$

Thus,

$$\mathbb{E}[\min(X,Y)] = \frac{n}{2} - \frac{1}{2}\mathbb{E}[|X - Y|] = \frac{n}{2} - \frac{1}{4}\mathbb{E}\left[\left|\sum_{i=1}^{2n} a_i\sigma_i\right|\right] \leq \frac{1}{2}n - \frac{1}{4}\sqrt{n} \ .$$

■