JMLR: Workshop and Conference Proceedings vol 35:1–26, 2014

Follow the Leader with Dropout Perturbations

Tim van Erven TIM@TIMVANERVEN.NL
Département de Mathématiques, Université Paris-Sud, France

Wojciech Kotłowski WKOTLOWSKI@CS.PUT.POZNAN.PL
Institute of Computing Science, Poznań University of Technology, Poland

Manfred K. Warmuth MANFRED@CSE.UCSC.EDU
Department of Computer Science, University of California, Santa Cruz

Abstract

We consider online prediction with expert advice. Over the course of many trials, the goal of the learning algorithm is to achieve small additional loss (i.e. regret) compared to the loss of the best from a set of K experts. The two most popular algorithms are Hedge/Weighted Majority and Follow the Perturbed Leader (FPL). The latter algorithm first perturbs the loss of each expert by independent additive noise drawn from a fixed distribution, and then predicts with the expert of minimum perturbed loss ("the leader") where ties are broken uniformly at random. To achieve the optimal worst-case regret as a function of the loss L^* of the best expert in hindsight, the two types of algorithms need to tune their learning rate or noise magnitude, respectively, as a function of L^*.

Instead of perturbing the losses of the experts with additive noise, we randomly set them to 0 or 1 before selecting the leader. We show that our perturbations are an instance of dropout — because experts may be interpreted as features — although for non-binary losses the dropout probability needs to be made dependent on the losses to get good regret bounds. We show that this simple, tuning-free version of the FPL algorithm achieves two feats: optimal worst-case $O(\sqrt{L^* \ln K} + \ln K)$ regret as a function of L^*, and optimal $O(\ln K)$ regret when the loss vectors are drawn i.i.d. from a fixed distribution and there is a gap between the expected loss of the best expert and all others.

A number of recent algorithms from the Hedge family (AdaHedge and FlipFlop) also achieve this, but they employ sophisticated tuning regimes. The dropout perturbation of the losses of the experts result in different noise distributions for each expert (because they depend on the expert's total loss) and curiously enough no additional tuning is needed: the choice of dropout probability only affects the constants.

Keywords: Online learning, regret bounds, expert setting, Follow the Leader, Follow the Perturbed Leader, Dropout

1. Introduction

We address the following online sequential prediction problem, known as *prediction with expert advice* (see e.g. Littlestone and Warmuth, 1994; Vovk, 1998): in each trial $t = 1, 2, \dots$ the algorithm (randomly) chooses one of K experts as its predictor before being told the loss vector $\boldsymbol{\ell}_t \in [0,1]^K$ of all experts. Let $L_{T,k}$ be the cumulative loss of the k-th expert after T rounds, and let $L^* = \min_k L_{T,k}$ be the cumulative loss of the best expert. Then the goal for the algorithm is to minimize the difference between its expected cumulative loss and L^*, which is known as the *regret*. In this paper, we consider not just the standard worst-case setting, in which losses are generated by an

© 2014 T. van Erven, W. Kotłowski & M.K. Warmuth.

adversary that tries to maximize the regret, but also a setting with easier data, in which loss vectors are sampled from a fixed probability distribution.

The simplest algorithm is to *Follow the Leader* (FTL), i.e. to choose an expert of minimal cumulative loss, with ties broken uniformly at random. This algorithm works well for probabilistic data, but unfortunately its regret will grow linearly with L^* and T on worst-case data, which means that it is not able to learn. This problem can be avoided by the Weighted Majority algorithm or its stratified version, the Hedge algorithm (Littlestone and Warmuth, 1994; Freund and Schapire, 1997), which chooses expert k with probability proportional to the *exponential weight* $\exp(-\eta L_{t-1,k})$ in trial t, for some positive *learning rate* η. For the appropriate tuning of η, Hedge is guaranteed to achieve sublinear regret of order $O(\sqrt{L^* \ln K} + \ln K)$, which is worst-case optimal (Cesa-Bianchi et al., 1997; Jiazhong et al., 2013).

In addition to using exponential weights, there is a second method that can achieve sublinear worst-case regret: perturb the cumulative losses of the experts by independent additive noise from a fixed distribution and then apply FTL to the perturbed losses. For exponentially distributed noise with its parameter appropriately chosen as a function of L^*, the resulting Follow the Perturbed Leader (FPL) algorithm can also guarantee regret $O(\sqrt{L^* \ln K} + \ln K)$ (Kalai and Vempala, 2005). FPL can also exactly simulate Hedge with any learning rate η by using log-exponential random variables for the noise that are scaled as a function of η (Kuzmin and Warmuth, 2005; Kalai, 2005).

Instead of using additive noise, we perturb the loss vector of the current trial: for binary losses $\boldsymbol{\ell}_t \in \{0,1\}^K$, we set each component $\ell_{t,k}$ to zero with fixed probability α. As we explain in Section 2, this is equivalent to *dropout* (Hinton et al., 2012), because $\ell_{t,k}$ may be interpreted as the k-th feature value at trial t. Surprisingly, this simple method achieves the same optimal regret bound for any value of $\alpha \in (0,1)$, without tuning.

We can also handle the general case when the loss vector $\boldsymbol{\ell}_t$ takes values in the continuous range $[0,1]^K$. In this case, we show that, on worst-case data, the vanilla dropout procedure that simply sets coordinates of the loss vector to zero, gets a suboptimal $\Omega(K)$ dependence on the number of experts K. However, by *binarizing* the dropout loss, by which we mean setting the loss $\ell_{t,k}$ of the k-th expert to 1 with probability $(1-\alpha)\ell_{t,k}$ and to 0 otherwise, FPL again achieves the optimal regret of $O(\sqrt{L^* \ln K} + \ln K)$.

Most Hedge or FPL algorithms depend on a learning rate or noise parameter $\eta > 0$, which needs to be tuned in terms of L^* to get the worst-case bound $O(\sqrt{L^* \ln K} + \ln K)$. A simple way is to maintain an estimate of L^* and tune η based on the current estimate. As soon as the loss of the best expert exceeds the current estimate, the estimate is doubled and the algorithm re-tuned (see e.g. Cesa-Bianchi et al., 1996, 1997). This method achieves the optimal regret for Hedge and the related previous versions of FPL, albeit with a large constant factor.

Much fancier tuning methods were introduced by Auer et al. (2002). Amongst others, their techniques have recently led to the AdaHedge and FlipFlop algorithms (van Erven et al., 2011; de Rooij et al., 2014), which work well in the case when the loss vector is drawn i.i.d. from a fixed distribution. In this case they achieve constant regret $O(\ln K)$ when there is a gap between the loss of the best expert and the expected loss of all others. Surprisingly, we can show that FPL based on the binarized dropout perturbations also achieves regret $O(\ln K)$ in the i.i.d. case, without the need of any tuning.

On the other side of the spectrum, there exists another simple perturbation based on *Random Walk Perturbation* (RWP) (Devroye et al., 2013): simply add an independent Bernoulli coin flip to each loss component in each trial (without any tuning). We highly benefited from the techniques used in the analysis of this version of FPL. However, we can show that RWP already has $\Omega(\sqrt{T})$ regret in the noise-free case $L^* = 0$, where T is the number of trials. In contrast, FPL with binarized dropout perturbation achieves constant regret of order $O(\ln K)$. It seems that since the perturbation of the loss of expert k depends on the total current loss of expert k, this makes FPL more adaptive and no additional tuning is required.

Our research opens up a large number of questions regarding tuning-free methods for learning with the linear loss.

1. Does FPL with dropout perturbations lead to efficient algorithms with close to optimal regret for shifting experts or for various combinatorial expert settings such as the online shortest path problem (Takimoto and Warmuth, 2003)? See (Devroye et al., 2013) for a similar discussion.

2. Are there other versions of dropout (such as dropping the entire loss vector with probability α) that achieve the same feats as the dropout perturbations used in this paper?

3. There is a natural matrix generalization of the expert setting (Warmuth and Kuzmin, 2011). Is there a version of dropout perturbation that avoids the expensive matrix decompositions used by all algorithms with good regret bounds for this generalization (Hazan et al., 2010) and replaces the matrix decompositions by maximum eigenvector calculations?

4. Wager et al. (2013) consider a Follow the Regularized Leader type algorithm for the batch setting. They use a regularizer that is a variation of the squared Euclidean distance, which they motivate as FTL on an approximation to the *expected* dropout loss. Although this use of dropout is entirely different from ours, it would be interesting to know whether it leads to novel regret or generalization bounds.

Outline The paper is organized as follows. In Section 2 we formally define dropout perturbation, and explain how dropping loss components $\ell_{t,k}$ is the same as dropping features. Section 3 contains the core of our paper: We prove that the regret of FPL with binarized dropout perturbations is bounded by $O(\sqrt{L^* \ln K} + \ln K)$. Even though our algorithm is simple to state, this proof is quite involved at this point. We construct a worst-case sequence of losses using a sequence of reductions. This sequence consists of two regimes and we prove bounds for each regime separately, which we then combine. We also clarify the difference to the RWP algorithm, which provably does not achieve this bound. Then, in Section 4, we show that dropout perturbation automatically adapts to i.i.d. loss vectors, for which it gets constant regret of order $O(\ln K)$. Some of the proof details are postponed to the appendix.

2. Dropout Perturbation

Setting Online prediction with expert advice (Cesa-Bianchi and Lugosi, 2006) is formalized as a repeated game between the algorithm and an adversary. At each iteration $t = 1, \ldots, T$, the algorithm randomly picks an expert $\hat{k}_t \in \{1, \ldots, K\}$ according to a probability distribution $\boldsymbol{w}_t = (w_{t,1}, \ldots, w_{t,K})$ of its choice. Then, the loss vector $\boldsymbol{\ell}_t \in [0,1]^K$ is revealed by the adversary, and the algorithm suffers loss $\ell_{t,\hat{k}_t}$. Let $L_{T,k} = \sum_{t=1}^{T} \ell_{t,k}$ be the cumulative loss of expert k. Then the goal of the algorithm is to minimize its *regret*, which is the difference between its expected

```
Algorithm 1: Follow the Leader with Binarized Dropout Perturbation (BDP)
Parameter     : Dropout probability $\alpha \in (0,1)$
Initialization: $\widetilde{L}_{0,k} = 0$ for all $k = 1, \dots, K$.
for $t = 1$ to $T$ do
    Pick expert $\hat{k}_t = \operatorname{argmin}_k \widetilde{L}_{t-1,k}$ (with ties broken uniformly at random).
    Observe loss vector $\boldsymbol{\ell}_t$ and suffer loss $\ell_{t,\hat{k}_t}$.
    Draw $\widetilde{\ell}_{t,k}$ according to (2.1), independently for all $k$.
    Update $\widetilde{L}_{t,k} = \widetilde{L}_{t-1,k} + \widetilde{\ell}_{t,k}$ for all $k$.
end
```

cumulative loss and the cumulative loss of the best expert chosen in hindsight:

$$\mathcal{R}_T = \mathbb{E}\left[\sum_{t=1}^{T} \ell_{t,\hat{k}_t}\right] - L^*, \qquad \text{where } L^* = \min_k L_{T,k}.$$

The expectation is taken with respect to the random choices of the algorithm, i.e. $\mathbb{E}\left[\sum_{t=1}^{T} \ell_{t,\hat{k}_t}\right] = \sum_{t=1}^{T} \boldsymbol{w}_t \cdot \boldsymbol{\ell}_t$. For the process that generates the losses, we will consider two different models: In the *worst-case* setting, studied in Section 3, we need to guarantee small regret for every possible sequence of losses $\boldsymbol{\ell}_1, \dots, \boldsymbol{\ell}_T \in [0,1]^K$, including the sequence that is the most difficult for the algorithm; in the *independent, identically distributed* (i.i.d.) setting, considered in Section 4, we assume the loss vectors $\boldsymbol{\ell}_t \in [0,1]^K$ all come from a fixed, but unknown probability distribution, which makes it possible to give stronger guarantees on the regret.

Follow the Perturbed Leader The general class of *Follow the Perturbed Leader* (FPL) algorithms (Kalai and Vempala, 2005; Hannan, 1957) is defined by the choice

$$\hat{k}_t = \operatorname*{argmin}_k \left\{ L_{t-1,k} + \xi_{t-1,k} \right\},$$

where $\xi_{t-1,k}$ is an additive random *perturbation* of the cumulative losses, which is chosen independently for every expert k. Kalai and Vempala (2005) consider exponential and uniform distributions for $\xi_{t-1,k}$ that depend on a parameter $\eta > 0$, and in the RWP algorithm of (Devroye et al., 2013) $\xi_{t-1,k} \sim \text{Binomial}(1/2, t) - t/2$ is a binomial variable on t outcomes with success probability $1/2$ that is centered around its mean.

Binarized Dropout Perturbations We introduce an instance of FPL based on *binarized dropout perturbation* (BDP). For any *dropout probability* $\alpha \in (0,1)$, the binarized dropout perturbation of loss $\ell_{t,k}$ is defined as:

$$\widetilde{\ell}_{t,k} = \begin{cases} 1 & \text{with probability } (1-\alpha)\ell_{t,k}, \\ 0 & \text{otherwise.} \end{cases} \tag{2.1}$$

Let $\widetilde{L}_{T,k} = \sum_{t=1}^{T} \widetilde{\ell}_{t,k}$ be the cumulative BDP loss for expert k. Then the BDP algorithm (see Algorithm 1) chooses

$$\hat{k}_t = \operatorname*{argmin}_k \widetilde{L}_{t-1,k},$$

with ties broken uniformly at random. BDP is conceptually very simple. Computationally, it might be of interest that it does *sparse updates*, which only operate on non-zero features: if $\ell_{t,k} = 0$, then $\widetilde{\ell}_{t,k} = 0$ as well, so, if one uses the same variable to store $\widetilde{L}_{t-1,k}$ and $\widetilde{L}_{t,k}$, then no update to the internal state of the algorithm is required. There is a parameter α, but this parameter only affects the constants in the theorems, so simply setting it to $\alpha = 1/2$ without any tuning already gives good performance. BDP may be viewed as an instance of FPL with the additive data-dependent perturbations

$$\xi_{t-1,k} = \widetilde{L}_{t-1,k} - L_{t-1,k}.$$

If the losses are binary, $\alpha = 1/2$ and the cumulative loss grows approximately as $L_{t-1,k} \approx t-1$ for all experts k, then $\xi_{t-1,k}$ is approximately distributed as $\text{Binomial}(1/2, t{-}1) - (t-1)$. Consequently, in this special case BDP is very similar to RWP because shifts by a constant do not change the minimizer $\hat{k}_t$.

Standard Dropout Perturbations Dropout is normally defined as dropping hidden units or features in a neural network while training the parameters of the network on batch data using Gradient Descent (Hinton et al., 2012; Wang and Manning, 2013; Wager et al., 2013). In each iteration, hidden units or features are dropped with a fixed probability. We are in the single neuron case, in which each expert may be identified with a feature and the *standard dropout* perturbations (without binarization) become[1]:

$$\widetilde{\ell}^{\mathrm{s}}_{t,k} = \begin{cases} \ell_{t,k} & \text{with probability } 1-\alpha, \\ 0 & \text{otherwise.} \end{cases}$$

For binary losses, standard dropout perturbation is the same as BDP, but for general losses in the interval $[0,1]$ they are different. We initially tried to prove our results for standard dropout perturbation, but it turns out that this approach cannot achieve the right dependence on the number of experts:

Theorem 2.1 *Consider the FPL algorithm based on standard dropout perturbation with parameter* $\alpha \in (0,1)$. *Then, for any* $B > 0$ *and any* $K \geq 2$, *there exists a loss sequence* $\boldsymbol{\ell}_1, \ldots, \boldsymbol{\ell}_T$ *with* $\boldsymbol{\ell}_t \in [0,1]^K$ *for which* $L^* \leq B$, *while the algorithm incurs regret at least* $\frac{1}{2}(K-1)$.

Proof Take ϵ and δ to be very small positive numbers specified later. The sequence is constructed as follows. First, all experts except expert 1 incur losses ϵ for a number of iterations that is sufficiently large that, with probability (w.p.) at least $1-\delta$, the algorithm will choose expert 1 as a leader. Note that the required number of iterations depends on δ, K and α, but it does not depend on ϵ. Then, expert 1 incurs unit loss (and so does the algorithm w.p. $1-\delta$), while all other experts incur no loss. Next, all experts except expert 2 incur losses ϵ for sufficiently many iterations that, w.p. at least $1-\delta$, the algorithm will choose expert 2 as a leader. Then, expert 2 incurs unit loss (and so does the algorithm w.p. $1-\delta$), while all other experts incur no loss. This process is then repeated for $k = 3, 4, \ldots, K-1$. At the end of the game, the algorithm incurred loss at least $K-1$ w.p.

[1] The loss $\boldsymbol{w} \cdot \boldsymbol{\ell}_t$ in round t that we consider here is *linear* in the weights $\boldsymbol{w}$. For a general convex loss of the form $f_t(\boldsymbol{w} \cdot \boldsymbol{x}_t)$, where $\boldsymbol{x}_t$ is the feature vector at round t, learning the weights $\boldsymbol{w}$ is often achieved by locally linearizing the loss with a first-order Taylor expansion: $f_t(\boldsymbol{w}_t \cdot \boldsymbol{x}_t) - f_t(\boldsymbol{w} \cdot \boldsymbol{x}_t) \leq \boldsymbol{w}_t \cdot \boldsymbol{\ell}_t - \boldsymbol{w} \cdot \boldsymbol{\ell}_t$ for the surrogate loss $\ell_{t,k} = f'_t(\boldsymbol{w}_t \cdot \boldsymbol{x}_t) x_{t,k}$. (See the discussion in Section 4.1 of Kivinen and Warmuth (1997) and Section 2.4 of Shalev-Shwartz (2011).) Thus, experts correspond to features, and setting the k-th feature $x_{t,k}$ to 0 is equivalent to setting $\ell_{t,k}$ to 0.

at least $1-(K-1)\delta$, while expert K incurred no more loss than $T\epsilon$. Taking δ small enough that $\delta < \frac{1}{2(K-1)}$, the expected loss of the algorithm is at least $\frac{1}{2}(K-1)$. Finally, we take ϵ small enough to satisfy $0 < T\epsilon \leq B$, which gives $L^* \leq B$. ■

For the case that $L^* \leq B \leq \ln K$, a better bound of order $O(\sqrt{L^* \ln K} + \ln K) = O(\ln K)$ is possible. So we see that standard dropout leads to a suboptimal algorithm, and therefore we use binarized dropout for the rest of the paper.

3. Regret Bounded in Terms of the Cumulative Loss of the Best Expert

We will prove that the regret for the BDP algorithm is bounded by $O(\sqrt{L^* \ln K} + \ln K)$:

Theorem 3.1 *For any sequence of losses taking values in* $[0,1]$ *with* $\min_k L_{T,k} = L^*$*, the regret of the BDP algorithm is bounded by*

$$\mathcal{R}_T \leq \frac{1}{\alpha^2(1-\alpha)}\Big(4\sqrt{2L^* \ln(3K)} + 3\ln(1+L^*)\Big) + \frac{\ln(3K)}{\alpha(1-\alpha)^2} + \frac{3}{\alpha}.$$

Since RWP is similar to BDP and is known to have regret bounded by $O(\sqrt{T \ln K})$, one might try to show that the regret for RWP satisfies the stronger bound $O(\sqrt{L^* \ln K} + \ln K)$ as well, but this turns out to be impossible:

Theorem 3.2 *For every* $T > 0$*, there exists a loss sequence* $\boldsymbol{\ell}_1, \ldots, \boldsymbol{\ell}_T$ *for which* $L^* = 0$*, while RWP suffers* $\Omega(\sqrt{T})$ *regret.*

This theorem, proved in Section A.1 of the appendix, shows that there is a fundamental difference between RWP and BDP, and that the data dependent perturbations used in BDP are crucial to obtaining a bound in terms of L^*.

We now turn to proving Theorem 3.1, starting with the observation that it is in fact sufficient to prove the theorem only for binary losses:

Lemma 3.3 *Let* $a \geq 0$ *and* b *be any constants. If the regret for the BDP algorithm is bounded by*

$$\mathcal{R}_T \leq a\sqrt{L^*} + b \tag{3.1}$$

for all sequences of binary losses, then it also satisfies (3.1) *for any sequence of losses with values in the whole interval* $[0,1]$.

Proof Let $\boldsymbol{\ell}_1, \ldots, \boldsymbol{\ell}_T$ be an arbitrary sequence of loss vectors with components $\ell_{t,k}$ taking values in the whole interval $[0,1]$. We need to show that the regret for this sequence satisfies (3.1). To this end, let $\boldsymbol{\ell}'_1, \ldots, \boldsymbol{\ell}'_T$ be an alternative sequence of losses that is generated randomly from $\boldsymbol{\ell}_1, \ldots, \boldsymbol{\ell}_T$ by letting

$$\ell'_{t,k} = \begin{cases} 0 & \text{with probability } 1-\ell_{t,k}, \\ 1 & \text{with probability } \ell_{t,k}, \end{cases}$$

independently for all t and k. Accordingly, let a prime on any quantity denote that it is evaluated on the alternative losses. For example, $L'_{T,k} = \sum_{t=1}^T \ell'_{t,k}$ is the cumulative alternative loss for expert k. Let $\boldsymbol{w}_t$ be the probability distribution on experts induced by the internal randomization of the

BDP algorithm, i.e. $w_{t,k} = P(\hat{k}_t = k)$, and let $\boldsymbol{w}'_t$ be its counterpart on the alternative losses. Then, because the BDP algorithm internally generates a binary sequence of losses with the same probabilities as $\ell'_{t,k}$, we have $\mathbb{E}_{\boldsymbol{\ell}'_1,\ldots,\boldsymbol{\ell}'_{t-1}}[\boldsymbol{w}'_t] = \boldsymbol{w}_t$, and, independently, $\mathbb{E}_{\boldsymbol{\ell}'_t}[\boldsymbol{\ell}'_t] = \boldsymbol{\ell}_t$. Consequently, $\mathbb{E}[\boldsymbol{w}'_t \cdot \boldsymbol{\ell}'_t] = \boldsymbol{w}_t \cdot \boldsymbol{\ell}_t$ and $\mathbb{E}[\widehat{L}'_T] = \widehat{L}_T$, where $\widehat{L}_T$ and $\widehat{L}'_T$ are the expected (with respect to the internal randomization) cumulative losses of the algorithm on the original and the alternative losses, respectively. Applying (3.1) for the alternative losses and taking expectations, it now follows by Jensen's inequality that

$$\begin{aligned}
\mathbb{E}[\widehat{L}'_T] - \mathbb{E}[\min_k L'_{T,k}] &\le a\,\mathbb{E}\left[\sqrt{\min_k L'_{T,k}}\right] + b \le a\sqrt{\mathbb{E}[\min_k L'_{T,k}]} + b \\
\mathbb{E}[\widehat{L}'_T] - \min_k \mathbb{E}[L'_{T,k}] &\le a\sqrt{\min_k \mathbb{E}[L'_{T,k}]} + b \\
\widehat{L}_T - \min_k L_{T,k} &\le a\sqrt{\min_k L_{T,k}} + b.
\end{aligned}$$

■

Thus, from now on, we assume the losses are binary, i.e. $\ell_{t,k} \in \{0,1\}$.

Following the standard approach for FPL algorithms of Kalai and Vempala (2005), we start by bounding the regret by applying the so-called *be-the-leader* lemma to the perturbed losses. Like in the analysis of Devroye et al. (2013), the result simplifies, because we can relate the expectation of the perturbed losses back to the original losses:

Lemma 3.4 (be-the-leader) *For any sequence of loss vectors $\boldsymbol{\ell}_1, \ldots, \boldsymbol{\ell}_T$, the regret for the BDP algorithm is bounded by*

$$\mathcal{R}_T \le \frac{1}{1-\alpha} \sum_{t=1}^{T} \mathbb{E}\left[\widetilde{\ell}_{t,\hat{k}_t} - \widetilde{\ell}_{t,\hat{k}_{t+1}}\right],$$

where $\widetilde{\boldsymbol{\ell}}_1, \ldots, \widetilde{\boldsymbol{\ell}}_T$ are the dropout perturbed losses and $\hat{k}_t$ is the random expert chosen by the BDP algorithm based on the perturbed losses before trial t.

Proof The expected regret is given by

$$\begin{aligned}
\mathcal{R}_T &= \sum_{t=1}^{T} \mathbb{E}\left[\ell_{t,\hat{k}_t}\right] - \min_k L_{T,k} = \sum_{t=1}^{T} \frac{1}{1-\alpha}\mathbb{E}\left[\widetilde{\ell}_{t,\hat{k}_t}\right] - \min_k \frac{1}{1-\alpha}\mathbb{E}\left[\widetilde{L}_{t,k}\right] \\
&\le \frac{1}{1-\alpha}\mathbb{E}\left[\sum_{t=1}^{T} \widetilde{\ell}_{t,\hat{k}_t} - \min_k \widetilde{L}_{T,k}\right] \le \frac{1}{1-\alpha}\mathbb{E}\left[\sum_{t=1}^{T} \left(\widetilde{\ell}_{t,\hat{k}_t} - \widetilde{\ell}_{t,\hat{k}_{t+1}}\right)\right],
\end{aligned}$$

where the second equality holds because $\hat{k}_t$ only depends on $\widetilde{\boldsymbol{\ell}}_1, \ldots, \widetilde{\boldsymbol{\ell}}_{t-1}$, which are independent of $\widetilde{\boldsymbol{\ell}}_t$, and $\mathbb{E}\left[\widetilde{\ell}_{t,k}\right] = (1-\alpha)\ell_{t,k}$; and the last inequality is from the be-the-leader lemma, applied to the perturbed losses, which shows that $\sum_{t=1}^{T} \widetilde{\ell}_{t,\hat{k}_{t+1}} \le \min_k \widetilde{L}_{T,k}$ (Kalai and Vempala, 2005, Equation 4), (Cesa-Bianchi and Lugosi, 2006, Lemma 3.1). ■

Every single term in the bound from Lemma 3.4 may be bounded further as follows:

$$\begin{aligned}
&\mathbb{E}\Big[\widetilde{\ell}_{t,\hat{k}_t} - \widetilde{\ell}_{t,\hat{k}_{t+1}}\Big] \\
&\quad = (1-\alpha)\sum_{k=1}^{K} \Pr(\hat{k}_t = k \neq \hat{k}_{t+1} \mid \widetilde{\ell}_{t,k} = \ell_{t,k})\ \mathbb{E}\left[\widetilde{\ell}_{t,\hat{k}_t} - \widetilde{\ell}_{t,\hat{k}_{t+1}} \mid \hat{k}_t = k \neq \hat{k}_{t+1}, \widetilde{\ell}_{t,k} = \ell_{t,k}\right] \\
&\quad \leq (1-\alpha)\sum_{k=1}^{K} \Pr(\hat{k}_t = k \neq \hat{k}_{t+1} \mid \widetilde{\ell}_{t,k} = \ell_{t,k})\ \ell_{t,k}. \qquad (3.2)
\end{aligned}$$

In the case considered by Kalai and Vempala (2005), this expression can be easily controlled, because, for their specific choice of perturbations, the conditional probability

$$P_{m,k} = \Pr(\hat{k}_{t+1} \neq k \mid \hat{k}_t = k, \widetilde{\ell}_{t,k} = \ell_{t,k}, \min_{j\neq k} \widetilde{L}_{t,k} = m)$$

is small, *uniformly for all* m (see the second display on p. 298 of their paper), which is easy to show, because $P_{m,k}$ depends only on the cumulative loss and the perturbations for a single expert k. Thus their choice of perturbations is what makes their analysis simple. Unfortunately, in our case and also for the RWP algorithm of Devroye et al. (2013), this simple approach breaks down, because, for the perturbations we consider, the probability $P_{m,k}$ may be large for some m (although such m will have small probability). We therefore cannot use a uniform bound on $P_{m,k}$, and as a consequence our proof is more complicated.

We also cannot follow the line of reasoning of Devroye et al., because it relies on the fact that, at any time t, the perturbation for every expert has a standard deviation of order $\sqrt{t}$. For our algorithm this is only true if the cumulative losses for the experts grow at least at a linear rate ct for some absolute constant $c > 0$, which need not be the case in general. Put differently, if the expert losses grow sublinearly, then our perturbations are too small to use the approach of Devroye et al..

Thus, we cannot use any of the existing approaches to control (3.2) for arbitrary losses, and, instead, we proceed as follows:

1. Given any integers $L_1, \ldots, L_K$, we find a canonical worst-case sequence of losses, which maximizes the regret of the BDP algorithm among all binary loss sequences of arbitrary length T such that $L_{T,k} = L_k$ for all experts k.

2. We bound (3.2) on this sequence.

3.1. The Canonical Worst-case Sequence

We will show that the worst-case regret is achieved for a sequence of unit vectors: let $\boldsymbol{u}_k \in \{0,1\}^K$ denote the *unit vector* $(0,\ldots,0,1,0,\ldots,0)$ where the 1 is in the k-th position. This restriction to unit vectors has been used before (Abernethy and Warmuth, 2010; Koolen and Warmuth, 2010), but here we go one step further. We build a canonical worst-case sequence of unit vectors from the following alternating schemes:

Definition 3.5 *For any* $k \in \{1,\ldots,K\}$, *we call a sequence of loss vectors* $\boldsymbol{u}_k, \ldots, \boldsymbol{u}_K$ *a* $k..K$-*alternating scheme.*

To simplify the notation in the proofs, we will henceforth assume that the experts are numbered from best to worst, i.e. $L_1 \leq \ldots \leq L_K$. In the canonical worst-case sequence, alternating schemes are repeated as follows:

Definition 3.6 (Canonical Worst-case Sequence) *Let $L_1 \leq \ldots \leq L_K$ be nonnegative integers. Among all sequences of binary losses of arbitrary length T such that $L_{T,k} = L_k$ for all k, we call the following sequence the* canonical worst-case sequence*:*

- *First repeat the $1..K$-alternating scheme L_1 times;*
- *then repeat the $2..K$-alternating scheme $L_2 - L_1$ times;*
- *then repeat the $3..K$-alternating scheme $L_3 - L_2$ times;*
- *and so on until finally we repeat the $K..K$-alternating scheme (which consists of just the unit vector $\boldsymbol{u}_K$) $L_K - L_{K-1}$ times.*

Note that the canonical worst-case sequence always consists of L_K alternating schemes. For example, the canonical worst-case sequence for cumulative losses $L_1 = 2, L_2 = 3, L_3 = 6$ consists of the following $L_3 = 6$ alternating schemes:

$$\underbrace{\begin{pmatrix}1\\0\\0\end{pmatrix},\begin{pmatrix}0\\1\\0\end{pmatrix},\begin{pmatrix}0\\0\\1\end{pmatrix}}_{\text{1..3-alternating scheme}},\underbrace{\begin{pmatrix}1\\0\\0\end{pmatrix},\begin{pmatrix}0\\1\\0\end{pmatrix},\begin{pmatrix}0\\0\\1\end{pmatrix}}_{\text{1..3-alternating scheme}},\underbrace{\begin{pmatrix}0\\1\\0\end{pmatrix},\begin{pmatrix}0\\0\\1\end{pmatrix}}_{\text{2..3-alternating scheme}},\underbrace{\begin{pmatrix}0\\0\\1\end{pmatrix}}_{\text{3..3-a.s.}},\underbrace{\begin{pmatrix}0\\0\\1\end{pmatrix}}_{\text{3..3-a.s.}},\underbrace{\begin{pmatrix}0\\0\\1\end{pmatrix}}_{\text{3..3-a.s.}}.$$

For the BDP algorithm, the following three operations do not increase the regret of a sequence of binary losses (see Section A.2 of the appendix for proofs):

1. If more than one expert gets loss in a single trial, then we split that trial into K trials in which we hand out the losses for the experts in turn (Lemma A.1).
2. If, in some trial, all experts get loss 0, then we remove that trial (Lemma A.2).
3. If, in two consecutive trials t and $t+1$, $\boldsymbol{\ell}_t = \boldsymbol{u}_k$ and $\boldsymbol{\ell}_{t+1} = \boldsymbol{u}_{k'}$ for some experts $k \neq k'$ such that $L_{t-1,k} \geq L_{t-1,k'}$, then we swap the trials, setting $\boldsymbol{\ell}_t = \boldsymbol{u}_{k'}$ and $\boldsymbol{\ell}_{t+1} = \boldsymbol{u}_k$ (Lemma A.3).

These operations can be used to turn any sequence of binary losses into the canonical worst-case sequence:

Lemma 3.7 *Let $L_1 \leq \ldots \leq L_K$ be nonnegative integers. Among all sequences of binary losses of arbitrary length T such that $L_{T,k} = L_k$ for all k, the regret of the BDP algorithm is (non-uniquely) maximized by its regret on the canonical worst-case sequence.*

Proof Start with an arbitrary sequence of binary losses such that $L_{T,k} = L_k$. By repeated application of operations 1 and 2, we turn this loss sequence into a sequence of unit vectors. We then repeatedly apply the swapping operation 3 to any trials t and $t+1$ such that $\boldsymbol{\ell}_t = \boldsymbol{u}_k$ and $\boldsymbol{\ell}_{t+1} = \boldsymbol{u}_{k'}$ for $k \neq k'$ where we have strict inequality $L_{t-1,k} > L_{t-1,k'}$. We do this until no such consecutive trials remain. This arrives at the canonical worst-case sequence except that the unit vectors within each alternating scheme may be permuted. However, by applying property 3 for the case $L_{t-1,k} = L_{t-1,k'}$, we can sort each alternating scheme into the canonical order. ■

Together with Lemma 3.3, the previous lemma implies that we just need to bound the regret of the BDP algorithm on the canonical worst-case loss sequence.

3.2. Regret on the Canonical Worst-case Sequence

By Lemma 3.4 and (3.2), the regret for the BDP algorithm is bounded by

$$\mathcal{R}_T \leq \sum_{t=1}^{T} \sum_{k=1}^{K} \Pr(\hat{k}_t = k \neq \hat{k}_{t+1} \mid \widetilde{\ell}_{t,k} = \ell_{t,k}) \ell_{t,k}. \tag{3.3}$$

On the canonical worst-case sequence, $\ell_{t,k}$ will be zero for all but a single expert k, which we can exploit to abbreviate notation: let $\hat{k}_t^+$ denote the leader (with ties still broken uniformly at random) on the perturbed cumulative losses if we would add 1 unit of loss to the perturbed cumulative loss $\widetilde{L}_{t-1,\hat{k}_t}$ of the expert selected by the BDP algorithm; and define the event

$$\mathcal{A}_t = \left\{\hat{k}_t = k \neq \hat{k}_t^+\right\} \qquad \text{for the expert } k \text{ such that } \ell_{t,k} = 1.$$

Then (3.3) simplifies to

$$\mathcal{R}_T \leq \sum_{t=1}^{T} \Pr(\mathcal{A}_t). \tag{3.4}$$

We split the $L_{T,K}$ alternating schemes of the canonical worst-case sequence into two regimes, and use different techniques to bound $\Pr(\mathcal{A}_t)$ in each case. Let $r \leq L_{T,K} - L_{T,1}$ be some nonnegative integer, which we will optimize later:

1. The first regime consists of the initial $L_{T,1} + r$ alternating schemes, where the difference in the cumulative loss between all experts is relatively small (at most r).

2. The second regime consists of the remaining $L_{T,K} - r - L_{T,1}$ alternating schemes. Now any expert that is still receiving losses will have a cumulative loss that is at least r larger than the cumulative loss of expert 1, and r will be chosen large enough to ensure that, with high probability, no such expert will the leader.

The First Regime

Define the *lead pack*

$$D_t = \{k \colon \widetilde{L}_{t-1,k} < \min_j \widetilde{L}_{t-1,j} + 2\},$$

which is the random set of experts that might potentially be the leader in trials t or $t+1$. As observed by Devroye et al. (2013), there can only be a leader change at time t if D_t contains more than one expert. Our goal will be to control the probability that this happens using the following lemma (proved in Section A.3.1), which is an adaptation of Lemma 3 of Devroye et al.:

Lemma 3.8 *Suppose $L_{t-1,k} \geq L > 0$ for all k. Then*

$$\Pr(|D_t| > 1) \leq \frac{1}{\alpha(1-\alpha)} \Big(\sqrt{\frac{2 \ln K}{L}} + \frac{3}{L}\Big).$$

Unfortunately, we cannot apply Lemma 3.8 immediately: in the first regime there are K trials for each of the first $L_{T,1}$ repetitions of the $1..K$-alternating scheme, and applying Lemma 3.8 to all of these trials would already give a bound of order $K\sqrt{L_{T,1} \ln K}$, which overshoots the right rate by a factor of K. To avoid this factor, we only apply the lemma once per alternating scheme, which is made possible by the following result:

Lemma 3.9 *Suppose an alternating scheme starts at trial s and ends at trial v. Then*

$$\sum_{t=s}^{v} \Pr(\mathcal{A}_t) \leq \frac{1}{\alpha}(v-s+1)\Pr(\mathcal{A}_{v+1}) \leq \frac{1}{\alpha}\Pr(\hat{k}_{v+1} \neq \hat{k}^{+}_{v+1}) \leq \frac{1}{\alpha}\Pr(|D_{v+1}| > 1).$$

Proof The first inequality follows from Lemmas A.4 and A.5 in Section A.3.2 of the appendix, which show that

$$\Pr(\mathcal{A}_s) \leq \Pr(\mathcal{A}_{s+1}) \leq \ldots \leq \Pr(\mathcal{A}_v) \leq \frac{1}{\alpha}\Pr(\mathcal{A}_{v+1}).$$

Let $k = K - (v - s)$ be the first expert in the alternating scheme. At time $v+1$ a new alternating scheme will start, so that the situation is entirely symmetrical between all experts that are part of the current alternating scheme:

$$\Pr(\mathcal{A}_{v+1}) = \Pr(\hat{k}_{v+1} = k' \neq \hat{k}^{+}_{v+1}) \qquad \text{for all } k' \in \{k, \ldots, K\}.$$

This implies the second inequality:

$$\begin{aligned}(v-s+1)\Pr(\mathcal{A}_{v+1}) &= \sum_{k'=k}^{K} \Pr(\hat{k}_{v+1} = k' \neq \hat{k}^{+}_{v+1}) \\ &= \Pr(\hat{k}_{v+1} \in \{k, \ldots, K\}, \hat{k}_{v+1} \neq \hat{k}^{+}_{v+1}) \leq \Pr(\hat{k}_{v+1} \neq \hat{k}^{+}_{v+1}).\end{aligned}$$

Finally, the last inequality follows by definition of the lead pack. ■

Applying either Lemma 3.8 or the trivial bound $\Pr(|D_t| > 1) \leq 1$ only to the trials $v+1$ that immediately follow an alternating scheme, we obtain the following result for the first regime of the canonical worst-case sequence (see Section A.3.3 for the proof):

Lemma 3.10 *Suppose the first regime ends with trial v. Then*

$$\sum_{t=1}^{v} \Pr(\mathcal{A}_t) \leq \frac{1}{\alpha}\left(\frac{1}{\alpha(1-\alpha)}\left(2\sqrt{2L_{T,1}\ln K} + 3\ln(1+L_{T,1})\right) + r + 1\right).$$

The Second Regime

We proceed to bound the probability of the event $\mathcal{A}_t$ during the second regime of the canonical worst-case sequence, using that

$$\Pr(\mathcal{A}_t) \leq \Pr(\hat{k}_t = k) \qquad \text{for the expert } k \text{ such that } \ell_{t,k} = 1.$$

This probability will be easy to control, because during the second regime the difference in cumulative loss between expert 1 and the experts that are still receiving losses, is sufficiently large that the BDP algorithm will prefer expert 1 over these experts with high probability.

Lemma 3.11 *Suppose the second regime starts in trial s. Then*

$$\sum_{t=s}^{T} \Pr(\mathcal{A}_t) \leq K\frac{2L_{T,1} + 3r}{2(1-\alpha)^2 r}\exp\left(\frac{-2(1-\alpha)^2 r^2}{2L_{T,1} + r}\right).$$

Proof Let t be a trial during the L-th alternating scheme, for some L in the second regime, and let k_t be the expert that gets loss in trial t. Then, by Hoeffding's inequality,

$$\begin{aligned}\Pr(\mathcal{A}_t) \leq \Pr(\hat{k}_t = k_t) \leq \Pr\left(\widetilde{L}_{t-1,k_t} \leq \widetilde{L}_{t-1,1}\right) &\leq \exp\left(\frac{-2(1-\alpha)^2(L-L_{T,1}-1)^2}{L+L_{T,1}-1}\right)\\ &= \exp\left(\frac{-2(1-\alpha)^2(L-L_{T,1}-1)^2}{2L_{T,1}+L-L_{T,1}-1}\right) \leq \exp\left(\frac{-2(1-\alpha)^2 r(L-L_{T,1}-1)}{2L_{T,1}+r}\right),\end{aligned}$$

where the last inequality follows from the fact that $\frac{x}{a+x}$ is increasing in x for $x \geq 0$ and $a > 0$. Let $s(L)$ and $v(L)$ denote the first and the last trial in the L-th alternating scheme. Then, summing up over all trials in the second regime, we get

$$\begin{aligned}\sum_{t=s}^{T}\Pr(\mathcal{A}_t) &= \sum_{L=L_{T,1}+r+1}^{L_{T,K}}\sum_{t=s(L)}^{v(L)}\Pr(\mathcal{A}_t) \leq \sum_{L=L_{T,1}+r+1}^{L_{T,K}}\sum_{t=s(L)}^{v(L)}\exp\left(\frac{-2(1-\alpha)^2 r(L-L_{T,1}-1)}{2L_{T,1}+r}\right)\\ &\leq K\sum_{L=L_{T,1}+r+1}^{L_{T,K}}\exp\left(\frac{-2(1-\alpha)^2 r(L-L_{T,1}-1)}{2L_{T,1}+r}\right) = K\sum_{x=r}^{L_{T,K}-L_{T,1}-1}\exp\left(\frac{-2(1-\alpha)^2 rx}{2L_{T,1}+r}\right)\\ &\leq K\exp\left(\frac{-2(1-\alpha)^2 r^2}{2L_{T,1}+r}\right) + K\sum_{x=r+1}^{\infty}\exp\left(\frac{-2(1-\alpha)^2 rx}{2L_{T,1}+r}\right).\end{aligned}$$

Here the second term is bounded by:

$$\int_r^{\infty}\exp\left(\frac{-2(1-\alpha)^2 rx}{2L_{T,1}+r}\right)\mathrm{d}x = \frac{2L_{T,1}+r}{2(1-\alpha)^2 r}\exp\left(\frac{-2(1-\alpha)^2 r^2}{2L_{T,1}+r}\right).$$

Putting things together we obtain

$$\begin{aligned}\sum_{t=s}^{T}\Pr(\mathcal{A}_t) &\leq K\exp\left(\frac{-2(1-\alpha)^2 r^2}{2L_{T,1}+r}\right) + K\frac{2L_{T,1}+r}{2(1-\alpha)^2 r}\exp\left(\frac{-2(1-\alpha)^2 r^2}{2L_{T,1}+r}\right)\\ &\leq K\frac{2L_{T,1}+3r}{2(1-\alpha)^2 r}\exp\left(\frac{-2(1-\alpha)^2 r^2}{2L_{T,1}+r}\right),\end{aligned}$$

which was to be shown. ■

The First and the Second Regime Together Combining the bounds for the first and second regimes from Lemmas 3.10 and 3.11, and optimizing r, we obtain the following result, which is proved in Section A.3.4 of the appendix:

Lemma 3.12 *On a canonical worst-case sequence of any length T for which $L_{T,1} = L^*$,*

$$\sum_{t=1}^{T}\Pr(\mathcal{A}_t) \leq \frac{1}{\alpha^2(1-\alpha)}\left(4\sqrt{2L^*\ln(3K)} + 3\ln(1+L^*)\right) + \frac{\ln(3K)}{\alpha(1-\alpha)^2} + \frac{3}{\alpha}.$$

Plugging this bound into (3.4) completes the proof of Theorem 3.1.

4. Constant Regret on IID Losses

In this section we show that our BDP algorithm has optimal $O(\ln K)$ regret when the loss vectors are drawn i.i.d. from a fixed distribution and there is a fixed gap γ between the expected loss of the best expert and all others.

Theorem 4.1 *Let $\gamma \in (0,1]$ and $\delta \in (0,1]$ be constants, and let k^* be a fixed expert. Suppose the loss vectors $\boldsymbol{\ell}_t$ are independent random variables such that the expected differences in loss satisfy*

$$\min_{k \neq k^*} \mathbb{E}[\ell_{t,k} - \ell_{t,k^*}] \geq \gamma \qquad \text{for all } t. \tag{4.1}$$

Then, with probability at least $1-\delta$, the regret of the BDP algorithm is bounded by a constant:

$$\mathcal{R}_T \leq \frac{8}{(1-\alpha)^2\gamma^2} \ln \frac{8K}{(1-\alpha)^2\gamma^2\delta} + 3. \tag{4.2}$$

As discussed in the proof in Section B.1, it is possible to improve the dependence on γ at the cost of getting a more complicated expression.

Acknowledgments

Tim van Erven was supported by NWO Rubicon grant 680-50-1112, Wojciech Kotłowski by the Foundation for Polish Science under the Homing Plus Program, and Manfred K. Warmuth by NSF grant IIS-1118028.

References

Jacob Abernethy and Manfred K. Warmuth. Repeated games against budgeted adversaries. In *Neural Information Processing Systems (NIPS)*, pages 1–9, 2010.

Peter Auer, Nicolò Cesa-Bianchi, and Claudio Gentile. Adaptive and self-confident on-line learning algorithms. *Journal of Computer and System Sciences*, 64(1):48–75, 2002.

Nicolò Cesa-Bianchi and Gábor Lugosi. *Prediction, learning, and games*. Cambridge University Press, 2006.

Nicolò Cesa-Bianchi, Philip M. Long, and Manfred K. Warmuth. Worst-case quadratic loss bounds for on-line prediction of linear functions by gradient descent. *IEEE Transactions on Neural Networks*, 7(2):604–619, May 1996.

Nicolò Cesa-Bianchi, Yaov Freund, David Haussler, David P. Helmbold, Robert E. Schapire, and Manfred K. Warmuth. How to use expert advice. *Journal of the ACM*, 44(3):427–485, 1997.

Steven de Rooij, Tim van Erven, Peter D. Grünwald, and Wouter M. Koolen. Follow the leader if you can, hedge if you must. *Journal of Machine Learning Research. To appear.*, 2014.

Luc Devroye, Gábor Lugosi, and Gergely Neu. Prediction by random-walk perturbation. In *Conference on Learning Theory (COLT)*, pages 460–473, 2013.

Yoav Freund and Robert E. Schapire. A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of Computer and System Sciences*, 55:119–139, 1997.

James Hannan. Approximation to Bayes risk in repeated play. *Contributions to the Theory of Games*, 3:97–139, 1957.

Elad Hazan, Satyen Kale, and Manfred K. Warmuth. On-line variance minimization in $O(n^2)$ per trial? In *Conference on Learning Theory (COLT)*, pages 314–315, 2010.

Geoffrey E. Hinton, Nitish Srivastava, Alex Krizhevsky, Ilya Sutskever, and Ruslan R. Salakhutdinov. Improving neural networks by preventing co-adaptation of feature detectors. *CoRR*, abs/1207.0580, 2012.

Nie Jiazhong, Wojciech Kotłowski, and Manfred K. Warmuth. On-line PCA with optimal regrets. In *Algorithmic Learning Theory (ALT)*, pages 98–112, 2013.

Adam Kalai. A perturbation that makes Follow the Leader equivalent to Randomized Weighted Majority. Private communication, December 2005.

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

Jyrki Kivinen and Manfred K. Warmuth. Additive versus Exponentiated Gradient updates for linear prediction. *Information and Computation*, 132(1):1–64, 1997.

Wouter M. Koolen and Manfred K. Warmuth. Hedging structured concepts. In *23rd Annual Conference on Learning Theory - COLT 2010*, pages 93–104. Omnipress, June 2010.

Dima Kuzmin and Manfred K. Warmuth. Optimum follow the leader algorithm. In *Conference on Learning Theory (COLT)*, pages 684–686, 2005. Open problem.

Nick Littlestone and Manfred K. Warmuth. The Weighted Majority algorithm. *Information and Computation*, 108(2):212–261, 1994.

Shai Shalev-Shwartz. Online learning and online convex optimization. *Foundations and Trends in Machine Learning*, 4(2):107–194, 2011.

Eiji Takimoto and Manfred K. Warmuth. Path kernels and multiplicative updates. *Journal of Machine Learning Research*, 4:773–818, 2003.

Tim van Erven, Peter D. Grünwald, Wouter Koolen, and Steven de Rooij. Adaptive hedge. In *Neural Information Processing Systems (NIPS)*, pages 1656–1664, 2011.

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

Stefan Wager, Sida Wang, and Percy Liang. Dropout training as adaptive regularization. In *Neural Information Processing Systems (NIPS)*, pages 351–359, 2013.

Sida I. Wang and Christopher D. Manning. Fast dropout training. In *International Conference on Machine Learning (ICML)*, pages 118–126, 2013.

Manfred K. Warmuth and Dima Kuzmin. Online variance minimization. *Machine Learning*, 87(1): 1–32, 2011.

Additional Material for “Follow the Leader with Dropout Perturbations”

Appendix A. Proof Details for Section 3

A.1. Proof of Theorem 3.2

Proof We prove the theorem by explicit construction of the sequence. We take $K = 2$ experts. The loss sequence is such that expert 1 never gets any loss (hence $L^* = L_{T,1} = 0$), while expert 2 gets losses $\epsilon_1, \epsilon_2, \dots, \epsilon_T$, where $\epsilon_t = \frac{1}{\sqrt{t}}$. This means that its cumulative loss $L_{t,2} = \sum_{s=1}^{t} \frac{1}{\sqrt{s}}$ is bounded by $2\sqrt{t+1} - 2 \leq L_{t,2} \leq 2\sqrt{t+1}$. We now bound the total loss of the RWP algorithm from below.

Consider the situation just before trial t. Then expert 2 has accumulated at most $L_{t-1,2} \leq 2\sqrt{t}$ loss so far. The expected loss of the algorithm in this iteration is ϵ_t times the probability that expert 1 becomes the leader. This probability is at least $P(\xi_{t-1,1} > \xi_{t-1,2} + 2\sqrt{t})$, where $\xi_{t,k} \sim \text{Binomial}(1/2, t) - t/2$. Consequently, $U_t = \xi_{t-1,1} - \xi_{t-1,2} + t$ is distributed as $\text{Binomial}(\frac{1}{2}, 2t)$. Therefore, the expected loss of the algorithm in iteration t is lower bounded by

$$\mathbb{E}\left[\ell_{t,\hat{k}_t}\right] \geq \epsilon_t P(U_t > 2\sqrt{t} + t) = \epsilon_t P\left(\frac{U_t - t}{\sqrt{t/2}} > 2\sqrt{2}\right).$$

We first give a rough idea of what happens. Due to Central Limit Theorem, the probability on the right hand side is approximately $P(Y > 2\sqrt{2})$, where $Y \sim N(0,1)$. Summing over trials we get that the expected cumulative loss of the algorithm is approximately $L_{T,2} \cdot P(Z > 2\sqrt{2}) = \Omega(\sqrt{T})$.

To be more precise, we use the Berry-Esséen theorem, which says that for $X \sim \text{Binomial}(n,p)$, and $Y \sim N(0,1)$, for any z,

$$\left| P\left(\frac{X - np}{\sqrt{np(1-p)}} > z\right) - P(Y > z) \right| \leq \frac{f(p)}{\sqrt{n}},$$

where $f(p)$ depends on p, but not on n. Using this fact with $n = 2t$ and $p = \frac{1}{2}$ results in

$$P\left(\frac{U_t - t}{\sqrt{t/2}} > 2\sqrt{2}\right) - P(Y > 2\sqrt{2}) \geq -\frac{c}{\sqrt{t}}$$

for some constant c. Therefore the expected cumulative loss of the algorithm can be lower bounded by

$$P(Y > 2\sqrt{2}) L_{T,2} - \sum_{t=1}^{T} \epsilon_t \frac{c}{\sqrt{t}} \geq P(Y > 2\sqrt{2}) L_{T,2} - c(\ln T + 1) = \Omega(\sqrt{T}),$$

which proves the theorem. ■

A.2. Operations to Reduce to the Canonical Worst-case Sequence

In this section we prove that the three operations on losses from Section 3.1 can only ever increase the regret of the BDP algorithm. As proved in Lemma 3.7, this implies that the canonical worst-case sequence maximizes the regret among all sequences of binary losses with the same cumulative losses $L_1, \dots, L_K$.

THE UNIT RULE HOLDS

We can assume without loss of generality that in every round exactly one expert gets loss. We call this the *unit rule*. It follows from two results, which will be proved in this subsection: first, Lemma A.1 shows that, if multiple experts get loss in the same trial, then the regret can only increase if we split that trial into multiple consecutive trials in which the experts get their losses in turn. Secondly, it is shown by Lemma A.2 that rounds in which all experts get zero loss do not change the regret, and can therefore be ignored in the analysis. (Although we only need them for binary losses, both lemmas hold for general losses with values in $[0, 1]$.)

Lemma A.1 (One Expert Gets Loss Per Trial) *Suppose $\boldsymbol{\ell}_1, \dots, \boldsymbol{\ell}_{t-1}, \boldsymbol{\ell}_t, \boldsymbol{\ell}_{t+1}, \dots, \boldsymbol{\ell}_T$ is a sequence of losses. Now consider the alternative sequence of losses $\boldsymbol{\ell}_1, \dots, \boldsymbol{\ell}_{t-1}, \boldsymbol{\ell}_t^1, \dots, \boldsymbol{\ell}_t^K, \boldsymbol{\ell}_{t+1}, \dots, \boldsymbol{\ell}_T$ with*

$$\ell_{t,k'}^k = \begin{cases} \ell_{t,k} & \text{if } k' = k, \\ 0 & \text{otherwise,} \end{cases}$$

for $k, k' = 1, \dots, K$. Then the regret of the BDP algorithm on the original losses $\mathcal{R}_T$ never exceeds its regret on the alternative losses $\mathcal{R}'_T$:

$$\mathcal{R}_T \leq \mathcal{R}'_T.$$

Proof The cumulative loss of the best expert L^* is the same on both sequences of losses, so we only need to consider the cumulative loss of the BDP algorithm. On trials $1, \dots, t-1$ and $t+1, \dots, T$ the algorithm's probability distributions on experts are the same for both sequences of losses, so we only need to compare its expected loss on $\boldsymbol{\ell}_t$ with its expected losses on $\boldsymbol{\ell}_t^1, \dots, \boldsymbol{\ell}_t^K$. To this end, we observe that the algorithm's probability of choosing expert k given losses $\boldsymbol{\ell}_1, \dots, \boldsymbol{\ell}_{t-1}$ is no more than its probability of choosing that expert given losses $\boldsymbol{\ell}_1, \dots, \boldsymbol{\ell}_{t-1}, \boldsymbol{\ell}_t^1, \dots, \boldsymbol{\ell}_t^{k-1}$, because during the trials $\boldsymbol{\ell}_t^1, \dots, \boldsymbol{\ell}_t^{k-1}$ expert k has received no loss whereas experts $1, \dots, k-1$ have respectively received losses $\ell_{t,1}, \dots, \ell_{t,k-1}$, which implies that the algorithm's expected loss can only increase on the alternative sequence of losses. ■

Lemma A.2 (Ignore All-zero Loss Vectors) *Suppose $\boldsymbol{\ell}_1, \dots, \boldsymbol{\ell}_{t-1}, \boldsymbol{\ell}_t, \boldsymbol{\ell}_{t+1}, \dots, \boldsymbol{\ell}_T$ is a sequence of losses such that $\boldsymbol{\ell}_{t,k} = 0$ for all k. Then the regret of the BDP algorithm is the same on the subsequence $\boldsymbol{\ell}_1, \dots, \boldsymbol{\ell}_{t-1}, \boldsymbol{\ell}_{t+1}, \dots, \boldsymbol{\ell}_T$ with trial t removed.*

Proof Both the best expert and the BDP algorithm have loss 0 on trial t. Trial t also does not influence the actions of the BDP algorithm for any trial $t' \neq t$. Hence the regret does not change if trial t is removed. ■

SWAPPING ARGUMENT

We call a loss vector $\boldsymbol{\ell}_t$ a *unit loss* if there exists a single k such that $\ell_{t,k} = 1$ while $\ell_{t,k'} = 0$ for all $k' \neq k$.

Let $\hat{k}_t$ denote the expert that is randomly selected by the BDP algorithm (based on some past losses $\boldsymbol{\ell}_1, \ldots, \boldsymbol{\ell}_{t-1}$) to predict trial t.

Let $\text{Binomial}(p, n)$ denote the distribution of a binomial random variable on n trials with success probability p.

Lemma A.3 (Swapping) *Suppose that $\boldsymbol{\ell}_1, \ldots, \boldsymbol{\ell}_{t-1}, \boldsymbol{\ell}_t, \boldsymbol{\ell}_{t+1}, \boldsymbol{\ell}_{t+2}, \ldots, \boldsymbol{\ell}_T$ is a sequence of unit losses with $\ell_{t,k} = \ell_{t+1,k'} = 1$ for $k \neq k'$ and that $L_{t-1,k'} \leq L_{t-1,k}$. Now consider the alternative sequence of losses $\boldsymbol{\ell}_1, \ldots, \boldsymbol{\ell}_{t-1}, \boldsymbol{\ell}_{t+1}, \boldsymbol{\ell}_t, \boldsymbol{\ell}_{t+2}, \ldots, \boldsymbol{\ell}_T$ in which $\boldsymbol{\ell}_t$ and $\boldsymbol{\ell}_{t+1}$ are swapped. Then the regret of the BDP algorithm on the original losses $\mathcal{R}_T$ never exceeds its regret on the alternative losses $\mathcal{R}'_T$:*

$$\mathcal{R}_T \leq \mathcal{R}'_T.$$

Proof The cumulative loss of the best expert L^* is the same on both sequences of losses, so we only need to consider the cumulative loss of the BDP algorithm. On trials $1, \ldots, t-1$ and $t+1, \ldots, T$ the algorithm's weights are the same for both sequences of loss, so we only need to compare its loss on $\boldsymbol{\ell}_t, \boldsymbol{\ell}_{t+1}$ in the original sequence with its loss on $\boldsymbol{\ell}_{t+1}, \boldsymbol{\ell}_t$ in the alternative sequence.

Let $\Pr$ and $\Pr'$ respectively denote probability with respect to the algorithm's randomness on the original and on the alternative losses. Since we assumed unit losses, we need to show that

$$\Pr(\hat{k}_t = k) + \Pr(\hat{k}_{t+1} = k') \leq \Pr{}'(\hat{k}_t = k') + \Pr{}'(\hat{k}_{t+1} = k). \tag{A.1}$$

As will be made precise below, we can express all these probabilities in terms of binomial random variables $\widetilde{L}_{t-1,j} \sim \text{Binomial}(1-\alpha, L_{t-1,j})$ for $j = 1, \ldots, K$ that represent the cumulative perturbed losses for the experts after $t-1$ trials, and an additional Bernoulli variable $X \sim \text{Binomial}(1-\alpha, 1)$ that represents whether the next loss is dropped out (depending on context, X is either $\widetilde{\ell}_{t,k}$ or $\widetilde{\ell}_{t,k'}$). It will also be convenient to define the derived values

$$M = \min_{j \neq k,k'} \widetilde{L}_{t-1,j}, \qquad C = |\{j \neq k, k' \colon \widetilde{L}_{t-1,j} = M\}|,$$

which represent the minimum perturbed loss and the number of experts achieving that minimum among all experts except k and k'. We will show that

$$\begin{aligned}\Pr(\hat{k}_t = k \mid M = m, C = c, X = x) + \Pr(\hat{k}_{t+1} = k' \mid M = m, C = c, X = x)\\ - \Pr{}'(\hat{k}_t = k' \mid M = m, C = c, X = x) - \Pr{}'(\hat{k}_{t+1} = k \mid M = m, C = c, X = x) \quad \text{(A.2)}\end{aligned}$$

is nonpositive for all m, c and x, from which (A.1) follows. In terms of the random variables we have defined, these conditional probabilities are

$$\begin{aligned}
\Pr(\hat{k}_t = k \mid M = m, C = c, X = x) &= \Pr(\widetilde{L}_k < \widetilde{L}_{k'}, \widetilde{L}_k < m) + \Pr(\widetilde{L}_k = \widetilde{L}_{k'} < m)\frac{1}{2} \\
&\quad + \Pr(\widetilde{L}_k = m < \widetilde{L}_{k'})\frac{1}{c+1} + \Pr(\widetilde{L}_k = m = \widetilde{L}_{k'})\frac{1}{c+2}, \\
\Pr(\hat{k}_{t+1} = k' \mid M = m, C = c, X = x) &= \Pr(\widetilde{L}_{k'} < \widetilde{L}_k + x, \widetilde{L}_{k'} < m) + \Pr(\widetilde{L}_{k'} = \widetilde{L}_k + x < m)\frac{1}{2} \\
&\quad + \Pr(\widetilde{L}_{k'} = m < \widetilde{L}_k + x)\frac{1}{c+1} + \Pr(\widetilde{L}_{k'} = m = \widetilde{L}_k + x)\frac{1}{c+2}, \\
\Pr{}'(\hat{k}_t = k' \mid M = m, C = c, X = x) &= \Pr(\widetilde{L}_{k'} < \widetilde{L}_k, \widetilde{L}_{k'} < m) + \Pr(\widetilde{L}_{k'} = \widetilde{L}_k < m)\frac{1}{2} \\
&\quad + \Pr(\widetilde{L}_{k'} = m < \widetilde{L}_k)\frac{1}{c+1} + \Pr(\widetilde{L}_{k'} = m = \widetilde{L}_k)\frac{1}{c+2}, \\
\Pr{}'(\hat{k}_{t+1} = k \mid M = m, C = c, X = x) &= \Pr(\widetilde{L}_k < \widetilde{L}_{k'} + x, \widetilde{L}_k < m) + \Pr(\widetilde{L}_k = \widetilde{L}_{k'} + x < m)\frac{1}{2} \\
&\quad + \Pr(\widetilde{L}_k = m < \widetilde{L}_{k'} + x)\frac{1}{c+1} + \Pr(\widetilde{L}_k = m = \widetilde{L}_{k'} + x)\frac{1}{c+2},
\end{aligned}$$

where for notational convenience, we skipped the subscript $t-1$ in $\widetilde{L}_{t-1,j}$ for all j. For $x = 0$, we see immediately that (A.2) is 0, so it remains only to consider the case $x = 1$. For that case, (A.2) simplifies to

$$\begin{aligned}
&\Pr(\widetilde{L}_k < \widetilde{L}_{k'}, \widetilde{L}_k < m) - \Pr(\widetilde{L}_k < \widetilde{L}_{k'} + 1, \widetilde{L}_k < m) \\
&\quad + \Pr(\widetilde{L}_{k'} < \widetilde{L}_k + 1, \widetilde{L}_{k'} < m) - \Pr(\widetilde{L}_{k'} < \widetilde{L}_k, \widetilde{L}_{k'} < m) \\
&\quad + \frac{1}{2}\Big(\Pr(\widetilde{L}_{k'} = \widetilde{L}_k + 1 < m) - \Pr(\widetilde{L}_k = \widetilde{L}_{k'} + 1 < m)\Big) \\
&\quad + \frac{1}{c+1}\Big(\Pr(\widetilde{L}_k = m < \widetilde{L}_{k'}) - \Pr(\widetilde{L}_k = m < \widetilde{L}_{k'} + 1)\Big) \\
&\quad + \frac{1}{c+1}\Big(\Pr(\widetilde{L}_{k'} = m < \widetilde{L}_k + 1) - \Pr(\widetilde{L}_{k'} = m < \widetilde{L}_k)\Big) \\
&\quad + \frac{1}{c+2}\Big(\Pr(\widetilde{L}_{k'} = m = \widetilde{L}_k + 1) - \Pr(\widetilde{L}_k = m = \widetilde{L}_{k'} + 1)\Big) \\
&= -\Pr(\widetilde{L}_k = \widetilde{L}_{k'} < m) + \Pr(\widetilde{L}_{k'} = \widetilde{L}_k < m) \\
&\quad + \frac{1}{2}\Big(\Pr(\widetilde{L}_{k'} = \widetilde{L}_k + 1 < m) - \Pr(\widetilde{L}_k = \widetilde{L}_{k'} + 1 < m)\Big) \\
&\quad - \frac{1}{c+1}\Big(\Pr(\widetilde{L}_k = m = \widetilde{L}_{k'}) + \Pr(\widetilde{L}_{k'} = m = \widetilde{L}_k)\Big) \\
&\quad + \frac{1}{c+2}\Big(\Pr(\widetilde{L}_{k'} = m = \widetilde{L}_k + 1) - \Pr(\widetilde{L}_k = m = \widetilde{L}_{k'} + 1)\Big) \\
&= \frac{1}{2}\sum_{a=1}^{m-1}\Big(\Pr(\widetilde{L}_{k'} = a = \widetilde{L}_k + 1) - \Pr(\widetilde{L}_k = a = \widetilde{L}_{k'} + 1)\Big) \\
&\quad + \frac{1}{c+2}\Big(\Pr(\widetilde{L}_{k'} = m = \widetilde{L}_k + 1) - \Pr(\widetilde{L}_k = m = \widetilde{L}_{k'} + 1)\Big).
\end{aligned}$$

To prove that this is nonpositive, it is sufficient to show that

$$\Pr(\widetilde{L}_{k'} = a = \widetilde{L}_k + 1) - \Pr(\widetilde{L}_k = a = \widetilde{L}_{k'} + 1) \leq 0 \tag{A.3}$$

for all nonnegative integers a. Abbreviate $L = L_{t-1,k}$ and $L' = L_{t-1,k'}$. If $a = 0$ or $a > L'$, then the left-most probability is 0 and (A.3) holds. Alternatively, for $a \in \{1, \ldots, L'\}$, the left-hand side of (A.3) is equal to

$$\left(\binom{L'}{a}\binom{L}{a-1} - \binom{L'}{a-1}\binom{L}{a}\right)(1-\alpha)^{2a-1}\alpha^{L+L'-2a+1},$$

so it is enough to show that

$$\binom{L'}{a}\binom{L}{a-1} - \binom{L'}{a-1}\binom{L}{a} \leq 0.$$

But this holds, because the left-hand side is equal to

$$\begin{aligned}
&\frac{L'!L!}{a!(L'-a)!(a-1)!(L-a+1)!} - \frac{L'!L!}{(a-1)!(L'-a+1)!a!(L-a)!} \\
&\qquad = \frac{L'!L!}{a!(L'-a+1)!(a-1)!(L-a+1)!}\Big((L'-a+1) - (L-a+1)\Big) \\
&\qquad\qquad = \frac{L'!L!}{a!(L'-a+1)!(a-1)!(L-a+1)!}\Big(L' - L\Big) \leq 0,
\end{aligned}$$

because $L' = L_{t-1,k'} \leq L_{t-1,k} = L$ by assumption. ■

A.3. Bounding Leader Changes on the Canonical Worst-case Sequence

A.3.1. PROOF OF LEMMA 3.8

Proof Within this proof, abbreviate $\widetilde{L}_k = \widetilde{L}_{t-1,k}$. Let $V_k \sim \text{Binomial}(1-\alpha, L)$ and $W_k \sim \text{Binomial}(1-\alpha, L_{t-1,k} - L)$, and assume $\widetilde{L}_k = V_k + W_k$. Also define $p(v) = \Pr(V_1 = v)$. Then

$$\begin{aligned}
\Pr(|D| = 1) &= \sum_{k=1}^{K} \Pr(\min_{j \neq k} \widetilde{L}_j \geq \widetilde{L}_k + 2) = \sum_{k=1}^{K}\sum_{v=0}^{L} p(v)\Pr(\min_{j \neq k} \widetilde{L}_j \geq v + W_k + 2) \\
&\geq \sum_{v=0}^{L-2}\sum_{k=1}^{K} p(v)\Pr(\min_{j \neq k} \widetilde{L}_j \geq v + W_k + 2) = \sum_{v=2}^{L}\sum_{k=1}^{K} p(v-2)\Pr(\min_{j \neq k} \widetilde{L}_j \geq v + W_k) \\
&= \sum_{v=2}^{L} \frac{p(v-2)}{p(v)} \underbrace{\sum_{k=1}^{K} p(v)\Pr(\min_{j \neq k} \widetilde{L}_j \geq v + W_k)}_{f(v)}.
\end{aligned} \tag{A.4}$$

Let $S = \{k\colon \widetilde{L}_k \leq \min_j \widetilde{L}_j\}$ be the set of leaders. Then

$$\begin{aligned} f(v) &= \sum_{k=1}^{K} \Pr(\min_{j \neq k} \widetilde{L}_j \geq \widetilde{L}_k, V_k = v) \geq \Pr(\exists k\colon \min_{j \neq k} \widetilde{L}_j \geq \widetilde{L}_k, V_k = v)\\ &= \Pr(\exists k\colon \min_{j} \widetilde{L}_j \geq \widetilde{L}_k, V_k = v) \geq \Pr(\min_{k \in S} V_k = v), \end{aligned}$$

where the first inequality follows by the union bound, and the second because the latter event implies the former. We remark that S is never empty, so that the minimum is always well-defined. We also have

$$\frac{p(v-2)}{p(v)} = \frac{\binom{L}{v-2}(1-\alpha)^{v-2}\alpha^{L-v+2}}{\binom{L}{v}(1-\alpha)^{v}\alpha^{L-v}} = \frac{\alpha^2}{(1-\alpha)^2} \cdot \frac{v(v-1)}{(L-v+2)(L-v+1)}.$$

Let $g(v) = \max\{\frac{v(v-1)}{(L-v+2)(L-v+1)}, 0\}$. Then, since $g(0) = g(1) = 0$, (A.4) is at least as large as

$$\frac{\alpha^2}{(1-\alpha)^2} \sum_{v=0}^{L} g(v) \Pr(\min_{k \in S} V_k = v) = \frac{\alpha^2}{(1-\alpha)^2} \mathbb{E}[g(V)]$$

for $V = \min_{k \in S} V_k$. The function $g(v)$ may be written as

$$g(v) = h_1(v) h_2(v) \qquad \text{for} \qquad h_1(v) = \frac{v}{L-v+2}, \quad h_2(v) = \frac{v-1}{L-v+1}.$$

For $v \geq 1$, both h_1 and h_2 are nonnegative, nondecreasing and convex, which implies that $g(v)$ also has these properties. Moreover, since $g(v) = 0$ for $v \in [0,1]$ all properties extend to all $v \geq 0$. Suppose that $\mathbb{E}[V] \geq (1-\alpha)L - B$ for some $B \geq 0$. Then Jensen's inequality and monotonicity of g imply that

$$\begin{aligned} \frac{\alpha^2}{(1-\alpha)^2} \mathbb{E}[g(V)] &\geq \frac{\alpha^2}{(1-\alpha)^2} g(\mathbb{E}[V]) \geq \frac{\alpha^2}{(1-\alpha)^2} g\Big((1-\alpha)L - B\Big)\\ &\geq \frac{\Big(\alpha L - \frac{\alpha}{1-\alpha} B\Big)\Big(\alpha L - \frac{\alpha}{1-\alpha}(B+1)\Big)}{(\alpha L + B + 2)(\alpha L + B + 1)} = \left(1 - \frac{\frac{1}{1-\alpha} B + 2}{\alpha L + B + 2}\right)\left(1 - \frac{\frac{1}{1-\alpha}(B+1)}{\alpha L + B + 1}\right)\\ &\geq 1 - \frac{\frac{1}{1-\alpha} B + 2}{\alpha L + B + 2} - \frac{\frac{1}{1-\alpha}(B+1)}{\alpha L + B + 1} \geq 1 - \frac{2B+3}{(1-\alpha)\alpha L}. \end{aligned}$$

Putting everything together, we find that

$$\Pr(|D_t| > 1) = 1 - \Pr(|D_t| = 1) \leq \frac{2B+3}{(1-\alpha)\alpha L}, \tag{A.5}$$

so that it remains to find a good bound B. Let $Y_k = L - V_k \sim \text{Binomial}(\alpha, L)$. Then

$$\begin{aligned} -\mathbb{E}[V] + (1-\alpha)L &= \mathbb{E}\left[\max_{k \in S}\big((1-\alpha)L - V_k\big)\right] \leq \mathbb{E}\left[\max_{k}\big((1-\alpha)L - V_k\big)\right]\\ &= \mathbb{E}\left[\max_{k}\big(Y_k - \alpha L\big)\right] \leq \sqrt{\frac{L \ln K}{2}} =: B, \end{aligned}$$

where the last inequality follows by a standard argument for sub-Gaussian random variables (see, for example, Lemmas A.13 and A.1 in the textbook by Cesa-Bianchi and Lugosi (2006)). Plugging this bound into (A.5) leads to the desired result. ■

A.3.2. PROOF DETAILS FOR LEMMA 3.9

Suppose that t is a trial during the first regime in which expert k gets a unit of loss. First we consider the case that k is not the last expert in a round of the alternating scheme:

Lemma A.4 *Suppose t is a trial during the first $L_{T,1} + r$ repetitions of the alternating scheme in which $\ell_{t,k} = 1$ for some $k < K$. Then*

$$\Pr(\mathcal{A}_t) \leq \Pr(\mathcal{A}_{t+1}).$$

Proof For $j \neq k, k+1$, let $\widetilde{L}_{t-1,j} \sim$ Binomial$(1-\alpha, L_{t-1,j})$ be the perturbed cumulative losses for all experts except experts k and $k+1$. Also define

$$M = \min_{j \neq k, k+1} \widetilde{L}_{t-1,j}, \qquad C = |\{j \neq k, k+1 \colon \widetilde{L}_{t-1,j} = M\}|.$$

By definition of the canonical worst-case sequence, expert k gets a unit of loss in trial t, expert $k+1$ will get a unit of loss in trial $t+1$, and $L_{t-1,k} = L_{t-1,k+1}$. We will construct the perturbed cumulative losses for experts k and $k+1$ from the following variables: $V, W \sim$ Binomial$(1-\alpha, L_{t-1,k})$ and $X \sim$ Binomial$(1-\alpha, 1)$. To express $\Pr(\mathcal{A}_t)$, we define the perturbed cumulative losses $\widetilde{L}_{t-1,k} = V$ and $\widetilde{L}_{t-1,k+1} = W$, but to express $\Pr(\mathcal{A}_{t+1})$ we let $\widetilde{L}_{t-1,k} = W + X$ and $\widetilde{L}_{t-1,k+1} = V$. This leads to

$$\begin{aligned}
&\Pr(\mathcal{A}_t \,|M = m, C = c) \\
&\quad = \Pr(V = m-1, W > m)\frac{c}{c+1} + \Pr(V = m-1, W = m)\frac{c+1}{c+2} \\
&\qquad + \Pr(V = m, W > m)\frac{1}{c+1} + \Pr(V = m, W = m)\frac{1}{c+2} \\
&\qquad + \Big(\Pr(V = W-1, W < m) + \Pr(V = W, W < m)\Big)\frac{1}{2}, \\
&\Pr(\mathcal{A}_{t+1} \,|M = m, C = c) \\
&\quad = \Pr(V = m-1, W+X > m)\frac{c}{c+1} + \Pr(V = m-1, W+X = m)\frac{c+1}{c+2} \\
&\qquad + \Pr(V = m, W+X > m)\frac{1}{c+1} + \Pr(V = m, W+X = m)\frac{1}{c+2} \\
&\qquad + \Big(\Pr(V = W+X-1, W+X < m) + \Pr(V = W+X, W+X < m)\Big)\frac{1}{2}
\end{aligned}$$

for any m and c. Thus

$$\begin{aligned}
&\Pr(\mathcal{A}_{t+1} \mid M = m, C = c) - \Pr(\mathcal{A}_t \mid M = m, C = c) \\
&\quad = \alpha\Big(\Pr(\mathcal{A}_{t+1} \mid M = m, C = c, X = 0) - \Pr(\mathcal{A}_t \mid M = m, C = c, X = 0)\Big) \\
&\qquad + (1-\alpha)\Big(\Pr(\mathcal{A}_{t+1} \mid M = m, C = c, X = 1) - \Pr(\mathcal{A}_t \mid M = m, C = c, X = 1)\Big) \\
&\quad = (1-\alpha)\Big(\Pr(\mathcal{A}_{t+1} \mid M = m, C = c, X = 1) - \Pr(\mathcal{A}_t \mid M = m, C = c)\Big), \qquad \text{(A.6)}
\end{aligned}$$

where

$$\begin{aligned}
&\Pr(\mathcal{A}_{t+1} \mid M = m, C = c, X = 1) - \Pr(\mathcal{A}_t \mid M = m, C = c)\\
&\quad= \Big(\Pr(V = m-1, W+1 > m) - \Pr(V = m-1, W > m)\Big)\frac{c}{c+1}\\
&\quad+ \Big(\Pr(V = m-1, W+1 = m) - \Pr(V = m-1, W = m)\Big)\frac{c+1}{c+2}\\
&\quad+ \Big(\Pr(V = m, W+1 > m) - \Pr(V = m, W > m)\Big)\frac{1}{c+1}\\
&\quad+ \Big(\Pr(V = m, W+1 = m) - \Pr(V = m, W = m)\Big)\frac{1}{c+2}\\
&\quad+ \Big(\Pr(V = W, W+1 < m) + \Pr(V = W+1, W+1 < m)\\
&\qquad - \Pr(V = W-1, W < m) - \Pr(V = W, W < m)\Big)\frac{1}{2}.
\end{aligned}$$

Using that V and W have the same distribution, so that we may switch their roles, this simplifies to

$$\begin{aligned}
&\Pr(\mathcal{A}_{t+1} \mid M = m, C = c, X = 1) - \Pr(\mathcal{A}_t \mid M = m, C = c)\\
&\quad= \Pr(V = m-1, W = m)\Big(\frac{c}{c+1} - \frac{c+1}{c+2} + \frac{1}{c+2}\Big)\\
&\qquad+ \Pr(V = W = m-1)\Big(\frac{c+1}{c+2} - \frac{1}{2}\Big)\\
&\qquad+ \Pr(V = W = m)\Big(\frac{1}{c+1} - \frac{1}{c+2}\Big)\\
&\quad\geq 0.
\end{aligned}$$

Substituting back into (A.6), we see that

$$\Pr(\mathcal{A}_{t+1} \mid M = m, C = c) - \Pr(\mathcal{A}_t \mid M = m, C = c) \geq 0 \qquad \text{for all } m \text{ and } c.$$

Hence $\Pr(\mathcal{A}_{t+1}) - \Pr(\mathcal{A}_t) \geq 0$ also holds unconditionally, from which the lemma follows. ■

Secondly, we consider the case that k is the last expert in a round of the alternating scheme:

Lemma A.5 *Suppose t is a trial during the first $L_{T,1} + r$ repetitions of the alternating scheme in which $\ell_{t,K} = 1$. Then*

$$\Pr(\mathcal{A}_t) \leq \frac{1}{\alpha}\Pr(\mathcal{A}_{t+1}).$$

(To make $\Pr(\mathcal{A}_{t+1})$ well-defined in case $t = T$, we adopt the convention that expert K gets a unit of loss in trial $T + 1$.)

Proof Let k be the expert that gets a unit of loss in trial $t + 1$ so that $\ell_{t+1,k} = 1$ and $\ell_{t+1,k'} = 0$ for $k' \neq k$. Trial $t + 1$ is at the beginning of a round of the alternating scheme, so by symmetry between the experts that are part of the alternating scheme, $\Pr(\mathcal{A}_{t+1})$ would remain the same if we changed $\boldsymbol{\ell}_{t+1}$ so that $\ell_{t+1,K} = 1$ and $\ell_{t+1,k'} = 0$ for all $k' \neq K$. But then we would have

$$\Pr(\mathcal{A}_{t+1} \mid \widetilde{\ell}_{t,K} = 0) = \Pr(\mathcal{A}_t),$$

where $\widetilde{\ell}_{t,K}$ is the perturbed loss for expert K in round t, and consequently

$$\begin{aligned}\Pr(\mathcal{A}_{t+1}) &= \alpha \Pr(\mathcal{A}_{t+1} \mid \widetilde{\ell}_{t,K} = 0) + (1-\alpha)\Pr(\mathcal{A}_{t+1} \mid \widetilde{\ell}_{t,K} = 1) \\ &\geq \alpha \Pr(\mathcal{A}_{t+1} \mid \widetilde{\ell}_{t,K} = 0) = \alpha \Pr(\mathcal{A}_t),\end{aligned}$$

from which the lemma follows. ■

A.3.3. PROOF OF LEMMA 3.10

Proof Let $v(L)$ denote the last trial in the L-th alternating scheme. Then, by Lemmas 3.9, 3.8 and the trivial bound $\Pr(|D_{v(L)+1}| > 1) \leq 1$,

$$\begin{aligned}\sum_{t=1}^{v} \Pr(\mathcal{A}_t) &\leq \frac{1}{\alpha} \sum_{L=1}^{L_{T,1}+r} \Pr(|D_{v(L)+1}| > 1) \leq \frac{1}{\alpha}\left(\frac{1}{\alpha(1-\alpha)} \sum_{L=2}^{L_{T,1}} \left(\sqrt{\frac{2\ln K}{L}} + \frac{3}{L}\right) + r + 1\right) \\ &\leq \frac{1}{\alpha}\left(\frac{1}{\alpha(1-\alpha)}\left(2\sqrt{2L_{T,1}\ln K} + 3\ln(1+L_{T,1})\right) + r + 1\right),\end{aligned}$$

where the last step follows from the fact that $\sum_{L=2}^{L_{T,1}} f(L) \leq \int_1^{L_{T,1}} f(L)\mathrm{d}L$ for any nonincreasing function f. ■

A.3.4. PROOF OF LEMMA 3.12

Here we show how to optimize r to obtain Lemma 3.12:
Proof Abbreviate $f(r) = \frac{2(1-\alpha)^2 r^2}{2L^*+r}$. Then the bound for the second regime from Lemma 3.11 can be written as

$$\sum_{t=s}^{T} \Pr(\mathcal{A}_t) \leq \frac{Kr}{f(r)} \cdot \frac{2L^*+3r}{2L^*+r} e^{-f(r)} \leq \frac{3Kr}{f(r)} e^{-f(r)}.$$

We will choose r to be the smallest nonnegative integer such that $f(r) \geq \ln(3K) \geq 1$, so that

$$\sum_{t=s}^{T} \Pr(\mathcal{A}_t) \leq r. \tag{A.7}$$

(It is no problem if this makes r exceed its maximal value $L_{T,K} - L_{T,1}$, because in that case the second regime is empty, so (A.7) still holds, and since the bound for the first regime from Lemma 3.10 is increasing in r it also still holds.)

To find r, we need to take the largest solution to

$$\frac{2(1-\alpha)^2 r^2}{2L^*+r} = \ln(3K),$$

and round it up to the nearest integer. Abbreviating $a = 2(1-\alpha)^2$ and $b = \ln(3K)$, this gives

$$r = \left\lceil \frac{b + \sqrt{b^2 + 8aL^*b}}{2a} \right\rceil \leq \frac{b}{a} + \sqrt{\frac{2L^*b}{a}} + 1 = \frac{\ln(3K)}{2(1-\alpha)^2} + \frac{\sqrt{L^*\ln(3K)}}{1-\alpha} + 1,$$

where the inequality follows from $\sqrt{x+y} \leq \sqrt{x} + \sqrt{y}$ for nonnegative x, y and $\lceil x \rceil \leq x+1$. Combining this bound on r with (A.7) and the bound for the first regime from Lemma 3.10, we find that

$$\begin{aligned}
\sum_{t=1}^{T} \Pr(\mathcal{A}_t) &\leq \frac{1}{\alpha}\left(\frac{1}{\alpha(1-\alpha)}\Big(2\sqrt{2L^* \ln K} + 3\ln(1+L^*)\Big) + r + 1\right) + r \\
&\leq \frac{1}{\alpha}\left(\frac{1}{\alpha(1-\alpha)}\Big(2\sqrt{2L^* \ln K} + 3\ln(1+L^*)\Big) + 2r + 1\right) \\
&\leq \frac{1}{\alpha}\left(\frac{1}{\alpha(1-\alpha)}\Big(2\sqrt{2L^* \ln K} + 3\ln(1+L^*)\Big) + \frac{\ln(3K)}{(1-\alpha)^2} + \frac{2\sqrt{L^* \ln(3K)}}{1-\alpha} + 3\right) \\
&\leq \frac{1}{\alpha}\left(\frac{1}{\alpha(1-\alpha)}\Big(4\sqrt{2L^* \ln(3K)} + 3\ln(1+L^*)\Big) + \frac{\ln(3K)}{(1-\alpha)^2} + 3\right),
\end{aligned}$$

which is equivalent to the statement of the lemma. ■

Appendix B. Proofs for Section 4

B.1. Proof of Theorem 4.1

Proof [Theorem 4.1] We will show that the conditions of Lemma B.1 below, with $c = \gamma/2$, are satisfied with probability at least $1-\delta$ if τ is chosen as

$$\tau = \left\lceil \frac{8}{(1-\alpha)^2\gamma^2} \ln \frac{8K}{(1-\alpha)^2\gamma^2\delta} \right\rceil .$$

Because the second term in the bound from Lemma B.1 is bounded by

$$\frac{4K}{(1-\alpha)^2\gamma^2} \exp\Big(-\frac{(1-\alpha)^2\gamma^2}{4}\tau\Big) \leq 1,$$

this shows that

$$\mathcal{R}_T \leq \mathcal{R}_{\tau+1} + 1$$

with probability at least $1-\delta$. We now get the inequality in (4.2) from the trivial bound $\mathcal{R}_{\tau+1} \leq \tau + 1$ and $\lceil x \rceil \leq x + 1$. Alternatively, one might also get a better dependence on γ by applying Theorem 3.1 to $\mathcal{R}_{\tau+1}$ and using that $L^* \leq \tau+1$, which leads to

$$\mathcal{R}_T = O\bigg(\frac{\ln K}{\gamma} + \frac{1}{\gamma}\sqrt{\ln\big(\tfrac{1}{\gamma^2\delta}\big)\ln K}\bigg).$$

To verify that the conditions of Lemma B.1 are satisfied with sufficient probability, define the following events for $t \geq \tau + 1$:

$$\mathcal{A}_{t,k}\colon L_{t,k} \geq \mathbb{E}[L_{t,k}] - \frac{\gamma}{4}t \quad \text{for } k \neq k^*, \qquad B_t\colon L_{t,k^*} \leq \mathbb{E}[L_{t,k^*}] + \frac{\gamma}{4}t,$$

and let $D_t = B_t \cap \bigcap_{k \neq k^*} \mathcal{A}_{t,k}$ be the event that they all hold simultaneously. By the assumption in (4.1) we have that

$$L_{t,k} - L_{t,k^*} \geq \mathbb{E}[L_{t,k}] - \mathbb{E}[L_{t,k^*}] - \frac{\gamma}{2}t \geq \frac{\gamma}{2}t \qquad \text{on } D_t,$$

so that the conditions of Lemma B.1 are satisfied with $c = \gamma/2$ if D_t holds for all $t \geq \tau + 1$.

By Hoeffding's inequality, the probabilities of the complementary events $\bar{A}_{t,k}$ and $\bar{B}_t$ are all bounded by $\exp(-\gamma^2 t/8)$ and hence by the union bound the probability of $\bar{D}_t$ is bounded by $K\exp(-\gamma^2 t/8)$. Combining this with another application of the union bound, we find that the probability that D_t fails to hold for any $t \geq \tau + 1$ is bounded by

$$\begin{aligned}\Pr\Big(\bigcup_{t=\tau+1,\ldots,T} \bar{D}_t\Big) &\leq \sum_{t=\tau+1}^{T} \Pr\left(\bar{D}_t\right) \leq K \sum_{t=\tau+1}^{T} \exp(-\frac{\gamma^2}{8}t) \\ &\leq K \int_{t=\tau}^{\infty} \exp(-\frac{\gamma^2}{8}t)\mathrm{d}t = \frac{8K}{\gamma^2}\exp(-\frac{\gamma^2}{8}\tau).\end{aligned}$$

The reader may verify that, for our choice of τ, this probability is bounded by δ, which completes the proof. ■

Lemma B.1 *Suppose that, for some k^*,*

$$L_{t,k} - L_{t,k^*} \geq ct \qquad \textit{for all } t \geq \tau + 1 \textit{ and } k \neq k^*,$$

where $c > 0$ is a constant, and τ is a nonnegative integer. Then the regret for the BDP algorithm is bounded by a constant:

$$\mathcal{R}_T \leq \mathcal{R}_{\tau+1} + \frac{K}{(1-\alpha)^2 c^2} \exp\Big(-(1-\alpha)^2 c^2 \tau\Big).$$

Proof From the assumption of the lemma we know that expert k^* will be the best expert at least for all $t \geq \tau + 1$. Consequently the regret is bounded by

$$\mathcal{R}_T \leq \mathcal{R}_{\tau+1} + \sum_{t=\tau+2}^{T} \Pr(\hat{k}_t \neq k^*).$$

For any $t \geq \tau + 2$ and any $k \neq k^*$, Hoeffding's inequality implies that

$$\begin{aligned}\Pr(\hat{k}_t = k) &\leq \Pr(\widetilde{L}_{t-1,k} \leq \widetilde{L}_{t-1,k^*}) \\ &= \Pr\Big(\widetilde{L}_{t-1,k^*} - \widetilde{L}_{t-1,k} - \mathbb{E}[\widetilde{L}_{t-1,k^*} - \widetilde{L}_{t-1,k}] \geq (1-\alpha)(L_{t-1,k} - L_{t-1,k^*})\Big) \\ &\leq \exp\Big(\frac{-2(1-\alpha)^2(L_{t-1,k} - L_{t-1,k^*})^2}{2(t-1)}\Big) \leq \exp\Big(-(1-\alpha)^2 c^2 (t-1)\Big).\end{aligned}$$

Consequently, by the union bound,

$$\begin{aligned}
\sum_{t=\tau+2}^{T} \Pr(\hat{k}_t \neq k^*) &\leq \sum_{t=\tau+2}^{T} \sum_{k \neq k^*} \Pr(\hat{k}_t = k) \leq (K-1) \sum_{t=\tau+2}^{T} \exp\Big(-(1-\alpha)^2 c^2 (t-1)\Big) \\
&= (K-1) \sum_{t=\tau+1}^{T-1} \exp\Big(-(1-\alpha)^2 c^2 t\Big) \leq (K-1) \int_{\tau}^{\infty} \exp\Big(-(1-\alpha)^2 c^2 t\Big) \mathrm{d}t \\
&\leq \frac{K}{(1-\alpha)^2 c^2} \exp\Big(-(1-\alpha)^2 c^2 \tau\Big),
\end{aligned}$$

from which the lemma follows. ■