Theoretical Computer Science 742 (2018) 50–65

Contents lists available at ScienceDirect

Theoretical Computer Science

www.elsevier.com/locate/tcs

On minimaxity of Follow the Leader strategy in the stochastic setting ☆

Wojciech Kotłowski [1]

Institute of Computing Science, Poznań University of Technology, Piotrowo 2, 60-965 Poznań, Poland

A R T I C L E   I N F O

Article history:
Received 30 January 2017
Received in revised form 15 September 2017
Accepted 9 October 2017
Available online 28 December 2017

Keywords:
Online learning
Prediction with expert advice
Follow the Leader
Risk
Regret
Pseudo-regret
Stochastic data
Minimax algorithm
Exponential families

A B S T R A C T

We consider the setting of prediction with expert advice with an additional assumption that each expert generates its losses i.i.d. according to some distribution. We first identify a class of "admissible" strategies, which we call permutation invariant, and show that every strategy outside this class will perform not better than some permutation invariant strategy. We then show that when the losses are binary, a simple Follow the Leader (FL) algorithm is the minimax strategy for this game, where minimaxity is simultaneously achieved for the expected regret, the pseudo-regret, and the excess risk. Furthermore, FL has also the smallest regret, pseudo-regret, and excess risk over all permutation invariant prediction strategies, simultaneously for all distributions over binary losses. We generalize these minimax results to the case in which each expert generates its losses from a distribution belonging to a one-dimensional exponential family, as well as to the case of loss vectors generated jointly from a multinomial distribution. We also show that when the losses are in the interval $[0, 1]$ and the learner competes against all distributions over $[0, 1]$, FL remains minimax only when an additional trick called "loss binarization" is applied.

1. Introduction

In the game of prediction with expert advice [2,3], the learner sequentially decides on one of K experts to follow, and suffers loss associated with the chosen expert. The difference between the learner's cumulative loss and the cumulative loss of the best expert is called *regret*. The goal is to minimize the regret in the worst case over all possible loss sequences. A prediction strategy which achieves this goal (i.e., minimizes the worst-case regret) is called *minimax*. While algorithms such as Weighted Majority/Hedge [4–6] or Follow the Perturbed Leader [7] guarantee the optimal worst-case regret in the asymptotic sense (i.e., their regrets grow at the optimal rate), there is no known exact solution to the minimax problem in the general setting. Still, it is possible to derive minimax algorithms for some special variants of this game: when the losses follow from evaluating binary predictions on binary labels [2,3], for binary losses with fixed loss budget [8], and when $K = 2$ [9].

☆ A preliminary version of this paper appeared at 27th International Conference on Algorithmic Learning Theory (ALT 2016) [1]. In this journal version we extended the background discussion and added two new sections on exponential family model (Section 4) and dependent experts (Section 6).

E-mail address: wkotlowski@cs.put.poznan.pl.

[1] The author was supported by the Polish National Science Centre under grant No. 2016/22/E/ST6/00299.

https://doi.org/10.1016/j.tcs.2017.12.030
0304-3975/© 2017 Published by Elsevier B.V.

Interestingly, all these minimax algorithms share a similar strategy of playing against a maximin adversary which assigns losses uniformly at random. They also have the *equalization* property: all data sequences lead to the same value of the regret. While this property makes them robust against the worst-case sequence, it also makes them over-conservative, preventing them from exploiting the case, when the actual data is not adversarially generated.

In this paper, we drop the analysis of worst-case performance entirely, and explore the minimax principle in a more constrained setting, in which the adversary is assumed to be *stochastic*. In particular, we associate with each expert k a fixed distribution P_k over loss values, and assume the observed losses of expert k are generated independently from P_k. The goal is then to determine the minimax algorithm under these stochastic assumptions. The motivation behind studying such a setting is in its practical usefulness: the data encountered in practice are rarely adversarial and can often be modeled as generated from a fixed (yet unknown) distribution (for instance, selecting the best classifier from a set of already trained candidates based on data gathered in an online manner).

We immediately face two difficulties here. First, due to stochastic nature of the adversary, it is no longer possible to follow standard approaches of minimax analysis, such as backward induction [2,3] or sequential minimax duality [10,9], and we need to resort to a different technique. We define the notion of *permutation invariance* of prediction strategies. This let us identify a class of "admissible" strategies (which we call permutation invariant), and show that every strategy outside this class will perform not better than some permutation invariant strategy. Secondly, while the regret is a single, commonly used performance metric in the worst-case setting, the situation is different in the stochastic case. We know at least three potentially useful metrics in the stochastic setting: the *expected regret*, the *pseudo-regret*, and the *excess risk* [11], and it is not clear, which of them should be used to define the minimax strategy.

Fortunately, it turns out that there exists a single strategy which is minimax with respect to all three metrics simultaneously. In the case of *binary* losses, which take out values from $\{0, 1\}$, this strategy turns out to be the *Follow the Leader* (FL) algorithm, which chooses an expert with the smallest cumulative loss at a given trial (with ties broken randomly). Interestingly, FL is known to perform poorly in the worst-case, as its worst-case regret will grow linearly with T [3]. On the contrary, in the stochastic setting with binary losses, FL has the smallest regret, pseudo-regret, and excess risk over all permutation invariant prediction strategies, *simultaneously for all distributions over binary losses!* We later show that all these minimax properties of FL strategy generalize to the case, in which each expert generates its losses from a distribution belonging to a one-dimensional exponential family (e.g., Gaussian, Bernoulli, Poisson, gamma, geometric, etc.), and the previously considered case of losses from $\{0, 1\}$ becomes a special case of the Bernoulli family.

Furthermore, we also show that the optimality of FL strategy breaks down in the case of losses in the range $[0, 1]$, in which each expert generates losses from an *arbitrary* distribution over $[0, 1]$. Here, FL is provably suboptimal. However, by applying *binarization trick* to the losses [12], i.e. randomly setting them to $\{0, 1\}$ such that the expectation matches the actual loss, and using FL on the binarized sequence (which results in the *binarized FL* strategy), we obtain the minimax strategy in this setup.

We finally consider the case of dependent experts, i.e. when the losses are i.i.d. between trials, but not necessarily between experts. While the general case turns out to be hard to approach, and our methods based on permutation invariance fail, we are able to analyze the simplest variant of dependent experts, where the loss vectors follow multinomial distribution, i.e. only a single expert gets loss in a given trial. We show that the FL strategy retains the minimax properties analogous to those given for binary losses and independent experts.

We note that when the excess risk is used as a performance metric, our setup falls into the framework of statistical decision theory [13,14], and the question we pose can be reduced to the problem of finding the minimax decision rule for a properly constructed loss function, which matches the excess risk on expectation. In principle, one could try to solve our problem by using the complete class theorem and search for the minimax rule within the class of (generalized) Bayesian decision rules. We initially followed this approach, but it turned out to be futile, as the classes of distributions we are considering are large (e.g., all distributions in the range $[0, 1]$), and exploring prior distributions over such classes becomes very difficult. On the other hand, the analysis presented in this paper is relatively simple, and works not only for the excess risk, but also for the expected regret and the pseudo-regret. To the best of our knowledge, both the results and the analysis presented here are novel.

We also note that there has recently been much work dedicated to combine almost optimal worst-case performance with good performance on "easy" (e.g., stochastic) sequences [15,12,16–18]. These methods, however, are motivated from different principles than the minimax principle, and their analysis is tangential to the topic of this work.

Follow the Leader strategy in the stochastic setting has already been analyzed extensively in the past. It is known that when the losses of experts are generated i.i.d., FL performs very well in terms of the expected regret [15,12,19]. Furthermore, the asymptotically optimal Upper-Confidence-Bound (UCB) algorithm used in the stochastic multi-armed bandit setting [3, 20] would reduce to FL in our ("full information") setup, as confidence intervals maintained by UCB for each expert would all be of the same size. If one uses excess risk as a performance metric, the setup considered here reduces to a simple scenario of learning in the finite hypothesis class in statistical learning theory, where it is known that Empirical Risk Minimization (equivalent to FL strategy) achieves $O(\sqrt{\log K/T})$ excess risk, which can be shown to be tight [21]. This immediately (by summing over trials) gives $O(\sqrt{T \log K})$ bound on the pseudo-regret of FL, and the same bound on the regret of FL, by using the fact that the difference between the pseudo-regret and the expected regret is independent of prediction strategy and lower than the expected regret of any online learning algorithm (which is, again, of order $O(\sqrt{T \log K})$). All these bounds hold even for dependent experts. In fact, since the tight lower bound $\Omega(\sqrt{T \log K})$ on the regret in the adversarial expert setup [2,3] is obtained by using *stochastic* adversary, it also applies in our setup and gives the rate of the minimax expected regret. However, what we show here is a stronger result than all of the above: FL (or its binarized version) is *exactly* the minimax strategy and no improvement can be made at any trial in the worst case.

The paper is organized as follows. In Section 2 we formally define the problem. The binary case is solved in Section 3, and extended to the exponential family model in Section 4. Section 5 concerns the case of all loss distributions over the interval $[0, 1]$, and Section 6 discusses the setup of dependent experts. Section 7 concludes the paper and presents some open problems.

2. Problem setting

2.1. Prediction with expert advice in the stochastic setting

In the game of prediction with expert advice, at each trial $t = 1, \ldots, T$, the learner predicts with a distribution $\boldsymbol{w}_t = (w_{t,1}, \ldots, w_{t,K})$ over K experts. Then, the loss vector $\boldsymbol{\ell}_t = (\ell_{t,1}, \ldots, \ell_{t,K}) \in \mathcal{X}^K$ is revealed ($\mathcal{X} \subseteq \mathbb{R}$ to be specified later), and the learner suffers loss:

$$\boldsymbol{w}_t \cdot \boldsymbol{\ell}_t = \sum_{k=1}^{K} w_{t,k}\ell_{t,k},$$

which can be interpreted as the expected loss the learner suffers by following one of the experts chosen randomly according to $\boldsymbol{w}_t$. Let $L_{t,k}$ denote the cumulative loss of expert k at the end of iteration t, $L_{t,k} = \sum_{q \leq t} \ell_{q,k}$. Let $\boldsymbol{\ell}^t$ abbreviate the sequence of losses $\boldsymbol{\ell}_1, \ldots, \boldsymbol{\ell}_t$. We will also use $\boldsymbol{\omega} = (\boldsymbol{w}_1, \ldots, \boldsymbol{w}_T)$ to denote the whole prediction strategy of the learner, having in mind that each distribution $\boldsymbol{w}_t$ is a function of the past $t-1$ outcomes $\boldsymbol{\ell}^{t-1}$.

In the worst-case (adversarial) formulation of the problem, the performance of prediction strategy $\boldsymbol{\omega}$ is measured by means of *regret*:

$$\sum_{t=1}^{T} \boldsymbol{w}_t \cdot \boldsymbol{\ell}_t - \min_k L_{T,k},$$

which is a difference between the algorithm's cumulative loss and the cumulative loss of the best expert. No assumption is made on the way the sequence of losses is generated, and hence the goal is to find an algorithm which minimizes the worst-case regret over all possible sequences $\boldsymbol{\ell}^T$.

In this paper, we drop the analysis of the worst-case performance and explore the minimax principle in the *stochastic* setting, defined as follows. We assume there are K distributions $\boldsymbol{P} = (P_1, \ldots, P_K)$ over $\mathcal{X}$, such that for each $k = 1, \ldots, K$, the losses $\ell_{t,k}$, $t = 1, \ldots, T$, are generated i.i.d. from P_k. Note that this implies that $\ell_{t,k}$ is independent from $\ell_{t',k'}$ whenever $t' \neq t$ or $k \neq k'$. The prediction strategy is then evaluated by means of the *expected regret*:

$$R_{\text{reg}}(\boldsymbol{\omega}, \boldsymbol{P}) = \mathbb{E}\left[\sum_{t=1}^{T} \boldsymbol{w}_t(\boldsymbol{\ell}^{t-1}) \cdot \boldsymbol{\ell}_t - \min_k L_{T,k}\right],$$

where the expectation is over loss sequences $\boldsymbol{\ell}^T$ with respect to distributions $\boldsymbol{P} = (P_1, \ldots, P_k)$, and we explicitly indicate the dependence of $\boldsymbol{w}_t$ on $\boldsymbol{\ell}^{t-1}$.

One may argue, however, that the expected regret does not properly capture the performance of a prediction strategy in the stochastic setting, as even with the full knowledge of distributions $\boldsymbol{P}$, the best the learner could do is to concentrate all mass of $\boldsymbol{w}_t$ on the expert which is best on expectation, i.e. $w_{t,k^*} = 1$ for any $k^* \in \operatorname{argmin}_k \mathbb{E}[L_{T,k}]$. Therefore, the expected regret cannot be reduced below $\min_k \mathbb{E}[L_{T,k}] - \mathbb{E}[\min_k L_{T,k}] \geq 0$ (which nonnegativity follows from Jensen's inequality). Thus, instead of comparing the algorithm's loss to the loss of the best expert on the actual outcomes, one can choose the *best expected* expert as a comparator, which leads to a metric:

$$R_{\text{pse}}(\boldsymbol{\omega}, \boldsymbol{P}) = \mathbb{E}\left[\sum_{t=1}^{T} \boldsymbol{w}_t(\boldsymbol{\ell}^{t-1}) \cdot \boldsymbol{\ell}_t\right] - \min_k \mathbb{E}\left[L_{T,k}\right],$$

which we call the *pseudo-regret* to be consistent with the multi-armed bandit literature [20], where a similar quantity is used.[2] Jensen's inequality implies $R_{\text{pse}}(\boldsymbol{\omega}, \boldsymbol{P}) \geq R_{\text{reg}}(\boldsymbol{\omega}, \boldsymbol{P})$ for any $\boldsymbol{\omega}$ and any $\boldsymbol{P}$, and the difference $R_{\text{pse}}(\boldsymbol{\omega}, \boldsymbol{P}) - R_{\text{reg}}(\boldsymbol{\omega}, \boldsymbol{P})$ is independent of $\boldsymbol{\omega}$ given fixed $\boldsymbol{P}$. This does not, however, imply that these metrics are equivalent in the minimax analysis, as the K-vector of distributions $\boldsymbol{P}$ is chosen by the adversary *against* strategy $\boldsymbol{\omega}$ played by learner, and this choice will

[2] A term *expected redundancy* is also used in information theory for a corresponding measure used to quantify the excess codelength of a prequential code [11].

Table 1
Performance measures.

Expected regret:	$R_{\rm reg}(\boldsymbol{\omega}, \boldsymbol{P}) = \mathbb{E}\left[\sum_{t=1}^{T} \boldsymbol{w}_t(\boldsymbol{\ell}^{t-1}) \cdot \boldsymbol{\ell}_t - \min_k L_{T,k}\right]$
Pseudo-regret:	$R_{\rm pse}(\boldsymbol{\omega}, \boldsymbol{P}) = \mathbb{E}\left[\sum_{t=1}^{T} \boldsymbol{w}_t(\boldsymbol{\ell}^{t-1}) \cdot \boldsymbol{\ell}_t\right] - \min_k \mathbb{E}\left[L_{T,k}\right]$
Excess risk:	$R_{\rm rsk}(\boldsymbol{\omega}, \boldsymbol{P}) = \mathbb{E}\left[\boldsymbol{w}_T(\boldsymbol{\ell}^{T-1}) \cdot \boldsymbol{\ell}_T\right] - \min_k \mathbb{E}\left[\ell_{T,k}\right]$

in general be different for the expected regret and the pseudo-regret. We also note that these two measures can differ drastically. For instance, when all experts generate their losses from the same Bernoulli distribution with parameter $\frac{1}{2}$, any algorithm has zero pseudo-regret, while the expected regret grows at the largest possible rate $\Omega(\sqrt{T \log K})$ [3].

Finally, the stochastic setting permits us to evaluate the prediction strategy by means of the *individual* rather than cumulative losses. Thus, it is reasonable to define the *excess risk* of a prediction strategy at time T:

$$R_{\rm rsk}(\boldsymbol{\omega}, \boldsymbol{P}) = \mathbb{E}\left[\boldsymbol{w}_T(\boldsymbol{\ell}^{T-1}) \cdot \boldsymbol{\ell}_T\right] - \min_k \mathbb{E}\left[\ell_{T,k}\right],$$

a metric traditionally used in statistics to measure the accuracy of statistical procedures.[3] Contrary to the expected regret and pseudo-regret defined by means of cumulative losses of the prediction strategy, the excess risk concerns only a single prediction at a given trial; hence, without loss of generality, we can choose the last trial T in the definition. For the sake of clarity, we summarize the three measures in Table 1.

Given performance measure R, we say that a strategy $\boldsymbol{\omega}^*$ is *minimax* with respect to R over the set of distributions $\mathcal{P}$, if:

$$\sup_{\boldsymbol{P} \in \mathcal{P}^K} R(\boldsymbol{\omega}^*, \boldsymbol{P}) = \inf_{\boldsymbol{\omega}} \sup_{\boldsymbol{P} \in \mathcal{P}^K} R(\boldsymbol{\omega}, \boldsymbol{P}),$$

where the infimum is over all prediction strategies, and the supremum is over all K-vectors of distributions $(P_1, \ldots, P_K)$, with $P_k \in \mathcal{P}$ for all $k = 1, \ldots, K$. In what follows, whenever $\mathcal{P}$ is the set of *all* distributions with the support on $\mathcal{X}$, and $\mathcal{X}$ is clear from the context, we use a shorthand notation $\sup_{\boldsymbol{P}}$ for $\sup_{\boldsymbol{P} \in \mathcal{P}^K}$.

2.2. Permutation invariance

In this section, we identify a class of "admissible" prediction strategies, which we call permutation invariant. The name comes from the fact that the performance of these strategies remains invariant under any permutation of the distributions $\boldsymbol{P} = (P_1, \ldots, P_K)$. We show that for every prediction strategy, there exists a corresponding permutation invariant strategy with not worse expected regret, pseudo-regret and excess risk in the worst-case with respect to all permutations of $\boldsymbol{P}$.

We say that a strategy $\boldsymbol{\omega}$ is *permutation invariant* if for any $t = 1, \ldots, T$, and any permutation $\sigma \in S_K$, where S_K denotes the group of permutations over $\{1, \ldots, K\}$, $\boldsymbol{w}_t(\sigma(\boldsymbol{\ell}^{t-1})) = \sigma(\boldsymbol{w}_t(\boldsymbol{\ell}^{t-1}))$, where for any vector $\boldsymbol{v} = (v_1, \ldots, v_K)$, we denote $\sigma(\boldsymbol{v}) = (v_{\sigma(1)}, \ldots, v_{\sigma(K)})$ and $\sigma(\boldsymbol{\ell}^{t-1}) = \sigma(\boldsymbol{\ell}_1), \ldots, \sigma(\boldsymbol{\ell}_{t-1})$. In words, if we σ-permute the indices of all past loss vectors, the resulting weight vector will be the σ-permutation of the original weight vector. Permutation invariant strategies are natural, as they only rely on the observed outcomes, not on the expert indices. We will show that lack of permutation invariance (e.g., when the strategy favors expert with a smaller index, etc.) can be exploited by the adversary to incur more loss to the learner. The performance of permutation invariant strategies remains the same under any permutation of the distributions from $\boldsymbol{P}$:

Lemma 1. *Let $\boldsymbol{\omega}$ be permutation invariant. Then, for any permutation $\sigma \in S_K$, $\mathbb{E}_{\sigma(\boldsymbol{P})}\left[\boldsymbol{w}_t(\boldsymbol{\ell}^{t-1}) \cdot \boldsymbol{\ell}_t\right] = \mathbb{E}_{\boldsymbol{P}}\left[\boldsymbol{w}_t(\boldsymbol{\ell}^{t-1}) \cdot \boldsymbol{\ell}_t\right]$, and moreover $R(\boldsymbol{\omega}, \sigma(\boldsymbol{P})) = R(\boldsymbol{\omega}, \boldsymbol{P})$, where R is the expected regret, pseudo-regret, or excess risk, and $\sigma(\boldsymbol{P}) = (P_{\sigma(1)}, \ldots, P_{\sigma(K)})$.*

Proof. We first show that the expected loss of the algorithm at any iteration $t = 1, \ldots, T$, is the same for both $\sigma(\boldsymbol{P})$ and $\boldsymbol{P}$:

$$\begin{aligned}
\mathbb{E}_{\sigma(\boldsymbol{P})}\left[\boldsymbol{w}_t(\boldsymbol{\ell}^{t-1}) \cdot \boldsymbol{\ell}_t\right] &= \mathbb{E}_{\boldsymbol{P}}\left[\boldsymbol{w}_t(\sigma^{-1}(\boldsymbol{\ell}^{t-1})) \cdot \sigma^{-1}(\boldsymbol{\ell}_t)\right] \\
&= \mathbb{E}_{\boldsymbol{P}}\left[\sigma^{-1}(\boldsymbol{w}_t(\boldsymbol{\ell}^{t-1})) \cdot \sigma^{-1}(\boldsymbol{\ell}_t)\right] \\
&= \mathbb{E}_{\boldsymbol{P}}\left[\boldsymbol{w}_t(\boldsymbol{\ell}^{t-1}) \cdot \boldsymbol{\ell}_t\right],
\end{aligned}$$

3 A similar metric is called *simple regret* in the multi-armed bandit literature [20].

where the first equality follows from $\boldsymbol{\ell}_t \sim \sigma(\boldsymbol{P}) \iff \sigma^{-1}(\boldsymbol{\ell}_t) \sim \boldsymbol{P}$, the second equality exploits the permutation invariance of $\boldsymbol{\omega}$, while the third equality uses a simple fact that the dot product is invariant under permuting both arguments. Therefore, the "loss of the algorithm" part of any of the three measures (regret, pseudo-regret, risk) remains the same. To show that the "loss of the best expert" part of each measure is the same, note that for any $t = 1, \ldots, T$, $k = 1, \ldots, K$, $\mathbb{E}_{\sigma(\boldsymbol{P})}\left[\ell_{t,k}\right] = \mathbb{E}_{\boldsymbol{P}}\left[\ell_{t,\sigma(k)}\right]$, which implies:

$$\begin{aligned}
\min_k \mathbb{E}_{\sigma(\boldsymbol{P})}\left[\ell_{T,k}\right] &= \min_k \mathbb{E}_{\boldsymbol{P}}\left[\ell_{T,\sigma(k)}\right] = \min_k \mathbb{E}_{\boldsymbol{P}}\left[\ell_{T,k}\right], \\
\min_k \mathbb{E}_{\sigma(\boldsymbol{P})}\left[L_{T,k}\right] &= \min_k \mathbb{E}_{\boldsymbol{P}}\left[L_{T,\sigma(k)}\right] = \min_k \mathbb{E}_{\boldsymbol{P}}\left[L_{T,k}\right], \\
\mathbb{E}_{\sigma(\boldsymbol{P})}\left[\min_k L_{T,k}\right] &= \mathbb{E}_{\boldsymbol{P}}\left[\min_k L_{T,\sigma(k)}\right] = \mathbb{E}_{\boldsymbol{P}}\left[\min_k L_{T,k}\right],
\end{aligned}$$

so that the "loss of the best expert" parts of all measures are also the same for both $\sigma(\boldsymbol{P})$ and $\boldsymbol{P}$. □

We now show that permutation invariant strategies are "admissible" in the following sense:

Theorem 2. *For any strategy $\boldsymbol{\omega}$, there exists permutation invariant strategy $\widetilde{\boldsymbol{\omega}}$, such that for any set of distributions $\boldsymbol{P}$,*

$$R(\widetilde{\boldsymbol{\omega}}, \boldsymbol{P}) = \max_{\sigma \in S_K} R(\widetilde{\boldsymbol{\omega}}, \sigma(\boldsymbol{P})) \le \max_{\sigma \in S_K} R(\boldsymbol{\omega}, \sigma(\boldsymbol{P})),$$

where R is either the expected regret, the pseudo-regret or the excess risk. In particular, this implies that: $\sup_{\boldsymbol{P} \in \mathcal{P}^K} R(\widetilde{\boldsymbol{\omega}}, \boldsymbol{P}) \le \sup_{\boldsymbol{P} \in \mathcal{P}^K} R(\boldsymbol{\omega}, \boldsymbol{P})$.

Proof. This first equality in the theorem immediately follows from Lemma 1. Define $\widetilde{\boldsymbol{\omega}} = (\widetilde{\boldsymbol{w}}_1, \ldots, \widetilde{\boldsymbol{w}}_T)$ as:

$$\widetilde{\boldsymbol{w}}_t(\boldsymbol{\ell}^{t-1}) = \frac{1}{K!} \sum_{\tau \in S_K} \tau^{-1}\Big(\boldsymbol{w}_t(\tau(\boldsymbol{\ell}^{t-1}))\Big).$$

Note that $\widetilde{\boldsymbol{\omega}}$ is a valid prediction strategy, since $\widetilde{\boldsymbol{w}}_t$ is a function of $\boldsymbol{\ell}^{t-1}$ and a distribution over K experts ($\widetilde{\boldsymbol{w}}_t$ is a convex combination of $K!$ distributions, so it is a distribution itself). Moreover, $\widetilde{\boldsymbol{\omega}}$ is permutation invariant:

$$\begin{aligned}
\widetilde{\boldsymbol{w}}_t(\sigma(\boldsymbol{\ell}^{t-1})) &= \frac{1}{K!} \sum_{\tau \in S_K} \tau^{-1}\Big(\boldsymbol{w}_t(\tau\sigma(\boldsymbol{\ell}^{t-1}))\Big) \\
&= \frac{1}{K!} \sum_{\tau \in S_K} (\tau\sigma^{-1})^{-1}\Big(\boldsymbol{w}_t(\tau(\boldsymbol{\ell}^{t-1}))\Big) \\
&= \frac{1}{K!} \sum_{\tau \in S_K} \sigma\tau^{-1}\Big(\boldsymbol{w}_t(\tau(\boldsymbol{\ell}^{t-1}))\Big) = \sigma(\widetilde{\boldsymbol{w}}_t(\boldsymbol{\ell}^{t-1})),
\end{aligned}$$

where the second equality is from replacing the summation index by $\tau \mapsto \tau\sigma$. Now, note that the expected loss of $\widetilde{\boldsymbol{w}}_t$ is:

$$\begin{aligned}
\mathbb{E}_{\boldsymbol{P}}\left[\widetilde{\boldsymbol{w}}_t(\boldsymbol{\ell}^{t-1}) \cdot \boldsymbol{\ell}_t\right] &= \frac{1}{K!} \sum_{\tau \in S_K} \mathbb{E}_{\boldsymbol{P}}\left[\tau^{-1}\Big(\boldsymbol{w}_t(\tau(\boldsymbol{\ell}^{t-1}))\Big) \cdot \boldsymbol{\ell}_t\right] \\
&= \frac{1}{K!} \sum_{\tau \in S_K} \mathbb{E}_{\boldsymbol{P}}\left[\boldsymbol{w}_t(\tau(\boldsymbol{\ell}^{t-1})) \cdot \tau(\boldsymbol{\ell}_t)\right] \\
&= \frac{1}{K!} \sum_{\tau \in S_K} \mathbb{E}_{\tau^{-1}(\boldsymbol{P})}\left[\boldsymbol{w}_t(\boldsymbol{\ell}^{t-1}) \cdot \boldsymbol{\ell}_t\right] \\
&= \frac{1}{K!} \sum_{\sigma \in S_K} \mathbb{E}_{\sigma(\boldsymbol{P})}\left[\boldsymbol{w}_t(\boldsymbol{\ell}^{t-1}) \cdot \boldsymbol{\ell}_t\right].
\end{aligned}$$

Since the "loss of the best expert" parts of all three measures are invariant under any permutation of $\boldsymbol{P}$ (see the proof of Lemma 1), we have:

$$R(\widetilde{\boldsymbol{\omega}}, \boldsymbol{P}) = \frac{1}{K!} \sum_{\sigma \in S_K} R(\boldsymbol{\omega}, \sigma(\boldsymbol{P})) \le \max_{\sigma \in S_K} R(\boldsymbol{\omega}, \sigma(\boldsymbol{P})). \tag{1}$$

This implies that:

$$\sup_{\boldsymbol{P}\in\mathcal{P}^K} R(\widetilde{\boldsymbol{\omega}}, \boldsymbol{P}) \leq \sup_{\boldsymbol{P}\in\mathcal{P}^K} \max_{\sigma\in S_K} R(\boldsymbol{\omega}, \sigma(\boldsymbol{P})) = \sup_{\boldsymbol{P}\in\mathcal{P}^K} R(\boldsymbol{\omega}, \boldsymbol{P}). \qquad \square$$

Theorem 2 states that strategies which are not permutation-invariant do not give any advantage over permutation-invariant strategies even when the set of distributions $\boldsymbol{P}$ is fixed (and even possibly known to the learner), but the adversary can permute the distributions to make the learner incur the most loss. We also note that one can easily show a slightly stronger version of Theorem 2: if strategy $\boldsymbol{\omega}$ is not permutation invariant, and it holds that $R(\boldsymbol{\omega}, \boldsymbol{P}) \neq R(\boldsymbol{\omega}, \tau(\boldsymbol{P}))$ for some set of distributions and some permutation τ, then $R(\widetilde{\boldsymbol{\omega}}, \boldsymbol{P}) < \max_{\sigma\in S_K} R(\boldsymbol{\omega}, \sigma(\boldsymbol{P}))$. This follows from the fact that the inequality in (1) becomes sharp.

2.3. Follow the Leader strategy

Given loss sequence $\boldsymbol{\ell}^{t-1}$, let $N_t = |\operatorname{argmin}_{j=1,\dots,K} L_{t-1,j}|$ be the size of the leader set at the beginning of trial t. We define the *Follow the Leader* (FL) strategy $\boldsymbol{w}_t^{\mathrm{fl}}$ such that $w_{t,k}^{\mathrm{fl}} = \frac{1}{N_t}$ if $k \in \operatorname{argmin}_j L_{t-1,j}$ and $w_{t,k}^{\mathrm{fl}} = 0$ otherwise. In other words, FL predicts with the current leader, breaking ties uniformly at random. It is straightforward to show that such defined FL strategy is permutation invariant.

3. Binary losses

In this section, we set $\mathcal{X} = \{0, 1\}$, so that all losses are binary, and let $\mathcal{P}$ be the set of all distributions on $\{0, 1\}$. In this case, each P_k is a Bernoulli distribution. Take any permutation invariant strategy $\boldsymbol{\omega}$. It follows from Lemma 1 that for any $\boldsymbol{P}$, and any permutation $\sigma \in S_K$, $\mathbb{E}_{\boldsymbol{P}}\left[\boldsymbol{w}_t(\boldsymbol{\ell}^{t-1}) \cdot \boldsymbol{\ell}_t\right] = \mathbb{E}_{\sigma(\boldsymbol{P})}\left[\boldsymbol{w}_t(\boldsymbol{\ell}^{t-1}) \cdot \boldsymbol{\ell}_t\right]$. Averaging this equality over all permutations $\sigma \in S_K$ gives:

$$\mathbb{E}_{\boldsymbol{P}}\left[\boldsymbol{w}_t(\boldsymbol{\ell}^{t-1}) \cdot \boldsymbol{\ell}_t\right] = \underbrace{\frac{1}{K!}\sum_{\sigma} \mathbb{E}_{\sigma(\boldsymbol{P})}\left[\boldsymbol{w}_t(\boldsymbol{\ell}^{t-1}) \cdot \boldsymbol{\ell}_t\right]}_{=:\, \overline{\mathrm{loss}}_t(\boldsymbol{w}_t, \boldsymbol{P})}, \tag{2}$$

where we defined $\overline{\mathrm{loss}}_t(\boldsymbol{w}_t, \boldsymbol{P})$ to be permutation-averaged expected loss at trial t. We now show the main result of this paper, a surprisingly strong property of FL strategy, which states that FL minimizes $\overline{\mathrm{loss}}_t(\boldsymbol{w}_t, \boldsymbol{P})$ *simultaneously* over all K-vectors of distributions. Hence, FL is not only optimal in the worst case, but is actually optimal for permutation-averaged expected loss *for any* $\boldsymbol{P}$, even if $\boldsymbol{P}$ is known to the learner! The consequence of this fact (by (2)) is that *FL has the smallest expected loss among all permutation invariant strategies for any* $\boldsymbol{P}$ (again, even if $\boldsymbol{P}$ is known to the learner).

Theorem 3. *Let $\boldsymbol{\omega}^{\mathrm{fl}} = (\boldsymbol{w}_1^{\mathrm{fl}}, \dots, \boldsymbol{w}_T^{\mathrm{fl}})$ be the FL strategy. Then, for any K-vector of distributions $\boldsymbol{P} = (P_1, \dots, P_K)$ over binary losses, for any strategy $\boldsymbol{\omega} = (\boldsymbol{w}_1, \dots, \boldsymbol{w}_T)$, and any $t = 1, \dots, T$:*

$$\overline{\mathrm{loss}}_t(\boldsymbol{w}_t^{\mathrm{fl}}, \boldsymbol{P}) \leq \overline{\mathrm{loss}}_t(\boldsymbol{w}_t, \boldsymbol{P}).$$

In particular, by (2), for any permutation-invariant strategy $\boldsymbol{\omega}$:

$$\mathbb{E}_{\boldsymbol{P}}\left[\boldsymbol{w}_t^{\mathrm{fl}}(\boldsymbol{\ell}^{t-1}) \cdot \boldsymbol{\ell}_t\right] \leq \mathbb{E}_{\boldsymbol{P}}\left[\boldsymbol{w}_t(\boldsymbol{\ell}^{t-1}) \cdot \boldsymbol{\ell}_t\right].$$

To prove the theorem, we need the following combinatorial result (proof of which is given in Appendix A), which can be interesting on its own right:

Lemma 4. *For any real numbers $x_1, \dots, x_K, y_1, \dots, y_K$, and any monotonically increasing function f, the expression:*

$$h(k) = \sum_{\sigma\in S_K} e^{\sum_{j=1}^K x_j f(y_{\sigma(j)})} y_{\sigma(k)}$$

is minimized by any $k^* \in \operatorname{argmin}_{j=1,\dots,K} x_j$.

Proof of Theorem 3. For any distribution P_k over binary losses, let $\mu_k := \mathbb{E}_{P_k}[\ell_{t,k}] = P_k(\ell_{t,k} = 1)$. We have:

$$\begin{aligned}\overline{\mathrm{loss}}_t(\boldsymbol{w}_t, \boldsymbol{P}) &= \frac{1}{K!}\sum_{\sigma} \mathbb{E}_{\sigma(\boldsymbol{P})}\left[\boldsymbol{w}_t(\boldsymbol{\ell}^{t-1}) \cdot \boldsymbol{\ell}_t\right] \\ &= \frac{1}{K!}\sum_{\sigma} \mathbb{E}_{\sigma(\boldsymbol{P})}\left[\boldsymbol{w}_t(\boldsymbol{\ell}^{t-1})\right] \cdot \mathbb{E}_{\sigma(\boldsymbol{P})}[\boldsymbol{\ell}_t]\end{aligned} \tag{3}$$

$$= \frac{1}{K!} \sum_{\sigma} \sum_{\boldsymbol{\ell}^{t-1}} \left(\prod_{k=1}^{K} \mu_{\sigma(k)}^{L_{t-1,k}} (1-\mu_{\sigma(k)})^{t-1-L_{t-1,k}} \right) \left(\sum_{k=1}^{K} w_{t,k}(\boldsymbol{\ell}^{t-1}) \mu_{\sigma(k)} \right)$$

$$= \frac{1}{K!} \sum_{\boldsymbol{\ell}^{t-1}} \underbrace{\sum_{k=1}^{K} w_{t,k}(\boldsymbol{\ell}^{t-1}) \left(\sum_{\sigma} \prod_{j=1}^{K} \mu_{\sigma(j)}^{L_{t-1,j}} (1-\mu_{\sigma(j)})^{t-1-L_{t-1,j}} \mu_{\sigma(k)} \right)}_{=: \, \overline{\mathrm{loss}}_t(\boldsymbol{w}_t, \boldsymbol{P} | \boldsymbol{\ell}^{t-1})},$$

where the sum indexed with $\boldsymbol{\ell}^{t-1}$ is over all loss sequences, and in the second equality we used the fact that $\boldsymbol{w}_t$ depends on $\boldsymbol{\ell}^{t-1}$ and does not depend on $\boldsymbol{\ell}_t$. Fix $\boldsymbol{\ell}^{t-1}$ and consider the term $\overline{\mathrm{loss}}_t(\boldsymbol{w}_t, \boldsymbol{P}|\boldsymbol{\ell}^{t-1})$. To finish the proof, it suffices to show that for any $\boldsymbol{P}$, and any $\boldsymbol{\ell}^{t-1}$, $\overline{\mathrm{loss}}_t(\boldsymbol{w}_t, \boldsymbol{P}|\boldsymbol{\ell}^{t-1})$ is minimized by setting $\boldsymbol{w}_t = \boldsymbol{e}_{k^*}$ for any $k^* \in \mathrm{argmin}_j L_{t-1,j}$, where $\boldsymbol{e}_k$ is the k-th standard basis vector with 1 on the k-th coordinate, and zeros on the remaining coordinates. Indeed, this implies that $\overline{\mathrm{loss}}_t(\boldsymbol{w}_t, \boldsymbol{P}|\boldsymbol{\ell}^{t-1})$ is also minimized by the FL strategy $\boldsymbol{w}_t^{\mathrm{fl}}$, which distributes its mass uniformly over all leaders. Then, by (3), it follows that $\boldsymbol{w}_t^{\mathrm{fl}}$ also minimizes $\overline{\mathrm{loss}}_t(\boldsymbol{w}_t, \boldsymbol{P})$.

Since $\overline{\mathrm{loss}}_t(\boldsymbol{w}_t, \boldsymbol{P}|\boldsymbol{\ell}^{t-1})$ is linear in $\boldsymbol{w}_t$, it suffices to check solutions of the form $\boldsymbol{w}_t = \boldsymbol{e}_k$, $k = 1, \ldots, K$, i.e. to show that for any $k^* \in \mathrm{argmin}_j L_{t-1,j}$ and any $k = 1, \ldots, K$:

$$\overline{\mathrm{loss}}_t(\boldsymbol{e}_{k^*}, \boldsymbol{P}|\boldsymbol{\ell}^{t-1}) \le \overline{\mathrm{loss}}_t(\boldsymbol{e}_k, \boldsymbol{P}|\boldsymbol{\ell}^{t-1}). \tag{4}$$

Assume for the moment that $\mu_j \notin \{0, 1\}$ for all j. We rewrite:

$$\overline{\mathrm{loss}}_t(\boldsymbol{e}_k, \boldsymbol{P}|\boldsymbol{\ell}^{t-1}) = \sum_{\sigma} \prod_{j=1}^{K} \mu_{\sigma(j)}^{L_{t-1,j}} (1-\mu_{\sigma(j)})^{t-1-L_{t-1,j}} \mu_{\sigma(k)}$$

$$= \left(\prod_{j=1}^{K} (1-\mu_j)^{t-1} \right) \sum_{\sigma} e^{\sum_{j=1}^{K} L_{t-1,j} \log \frac{\mu_{\sigma(j)}}{1-\mu_{\sigma(j)}}} \mu_{\sigma(k)}.$$

Then, (4) is implied by Lemma 4 with $x_j = L_{t-1,j}$, $y_j = \mu_j$, $j = 1, \ldots, K$, and $f(y) = \log \frac{y}{1-y}$, which is monotonically increasing. To account for the case $\mu_j \in \{0, 1\}$ note, that since (4) holds for $\mu_j = \epsilon$ (or $\mu_j = 1 - \epsilon$) with arbitrarily small $\epsilon > 0$, it must also hold for $\mu_j = 0$ (or $\mu_j = 1$) by taking the limit $\epsilon \to 0$ and using the continuity of $\overline{\mathrm{loss}}_t(\boldsymbol{w}_t, \boldsymbol{P}|\boldsymbol{\ell}^{t-1})$ with respect to $\boldsymbol{P}$. □

Note that the proof did not require uniform tie breaking over leaders, as any distribution over leaders would work as well. Uniform distribution, however, makes the FL strategy permutation invariant.

The consequence of Theorem 3 is the following corollary which states the minimaxity of FL strategy for binary losses:

Corollary 5. *Let $\boldsymbol{\omega}^{\mathrm{fl}} = (\boldsymbol{w}_1^{\mathrm{fl}}, \ldots, \boldsymbol{w}_T^{\mathrm{fl}})$ be the FL strategy. Then, for any $\boldsymbol{P}$ over binary losses, and any permutation invariant strategy $\boldsymbol{\omega}$:*

$$R(\boldsymbol{\omega}^{\mathrm{fl}}, \boldsymbol{P}) \le R(\boldsymbol{\omega}, \boldsymbol{P}),$$

where R is the expected regret, pseudo-regret, or excess risk. This implies:

$$\sup_{\boldsymbol{P}} R(\boldsymbol{\omega}^{\mathrm{fl}}, \boldsymbol{P}) = \inf_{\boldsymbol{\omega}} \sup_{\boldsymbol{P}} R(\boldsymbol{\omega}, \boldsymbol{P}),$$

where the supremum is over all distributions on binary losses, and the infimum over all (not necessarily permutation invariant) strategies.

Proof. The first statement follows from the second statement of Theorem 3, and the fact that the “loss of the best expert” part is the same of on both sides on inequality.

The second statement immediately follows from the first statement and Theorem 2. □

4. Exponential family model

So far, we discussed a setting in which the losses generated by each expert are in $\{0, 1\}$. Here, we extend the minimax analysis to the case where each expert is a member of a one-dimensional exponential family of distributions.

Let $\mathcal{X} \subseteq \mathbb{R}$ be a set of outcomes, which can be a finite or countable set, or a general subset of the reals. One-dimensional exponential family [22,11] with *sufficient statistic* $\phi : \mathcal{X} \to \mathbb{R}$ and *carrier* $h : \mathcal{X} \to [0, \infty)$ is the family of distributions on $\mathcal{X}$ with densities:

$$P_\theta(x) = h(x)e^{\theta\phi(x)-\psi(\theta)}, \qquad \theta \in \Theta.$$

Here $\psi(\theta) = \log \int_{x\in\mathcal{X}} e^{\theta\phi(x)}h(x)\,\mathrm{d}x$ (with the integral replaced by a sum for countable $\mathcal{X}$) is the cumulant generating function, and $\Theta = \{\theta \in \mathbb{R} : \psi(\theta) < \infty\}$ is the *natural parameter space*. Examples of exponential families include Gaussian distribution with fixed variance *or* fixed mean, Bernoulli, Poisson, exponential, gamma (with one of the parameters fixed), geometric, Pareto families of distribution, and many others. We will simplify the presentation and assume that $\phi(x) \equiv x$, i.e. the exponential family is in the canonical form. All our results are valid for a more general ϕ (in this case, however, one should replace losses $\ell_{t,k}$ by $\phi(\ell_{t,k})$ for all t, k).

A standard result for exponential families states [22] that $\psi(\theta)$ is differentiable infinitely often, strictly convex on Θ, and the derivative $\mu(\theta) = \psi'(\theta)$ is the mean value of the sufficient statistics, $\mu(\theta) = \mathbb{E}_\theta[x]$. Due to strict convexity of ψ, $\mu(\theta)$ is strictly increasing, so that there is one-to-one mapping between Θ and its image $M = \mu(\Theta)$, which is called the *mean-value parameter space*. We let $\theta(\mu)$ denote the inverse of $\mu(\theta)$, which maps parameters in the mean-value parameterization back to the natural parameterization.

Let $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$ be a one-dimensional exponential family. We associate with each expert $k = 1, \dots, K$ a natural parameter $\theta_k \in \Theta$ (or, equivalently, a mean-value parameter $\mu_k = \mu(\theta_k) \in M$), which is unknown to the learner, and the losses of expert k, $\ell_{t,k}$, $t = 1, \dots, T$ are generated i.i.d. from P_{θ_k}. Note that $\mathbb{E}_{P_{\theta_k}}[\ell_{t,k}] = \mu_k$ from the definition.

Example. Let $\mathcal{P}$ be the family of Bernoulli distributions, i.e. the family of all distributions on $\mathcal{X} = \{0, 1\}$:

$$P_\mu(x) = \mu^x(1-\mu)^{1-x} \quad \text{(mean-value)}, \qquad P_\mu(x) = e^{x\theta - \log(1+e^\theta)} \quad \text{(natural)}.$$

For a given expert k, $\mu_k = P(\ell_{t,k} = 1)$ is its mean value parameter, while $\theta_k = \log \frac{\mu_k}{1-\mu_k}$ is the corresponding natural parameter. Thus, the case of binary losses analyzed in Section 3 is a special case of exponential family model if one chooses the exponential family to be the family of Bernoulli distributions.

Example. Let $\mathcal{P}$ be the family of one-dimensional Gaussian distributions with fixed variance σ^2:

$$P_\mu(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \text{ (mean-value)}, \qquad P_\theta(x) = \frac{e^{-\frac{x^2}{2\sigma^2}}}{\sqrt{2\pi\sigma^2}} e^{\theta x - \frac{\theta^2\sigma^2}{2}} \text{ (natural)}.$$

Here the mean-value and natural parameterizations essentially coincide as we have $\mu = \theta\sigma^2$. Note that $\mathcal{X} = \mathbb{R}$, so that the range of losses is unbounded.

Example. Let $\mathcal{P}$ be the family of Poisson distributions:

$$P_\mu(x) = \frac{\mu^x e^{-\mu}}{x!} \quad \text{(mean-value)}, \qquad P_\theta(x) = \frac{1}{x!} e^{x\theta - e^\theta} \quad \text{(natural)}.$$

Here, $\mathcal{X} = \{0, 1, 2, \dots\}$, and $\theta = \log \mu$.

We now show that in the exponential family model, the FL strategy admits a property analogous to given in Theorem 3 for binary losses, i.e. that FL strategy minimizes the permutation-averaged loss simultaneously over all K-vectors of distributions $\boldsymbol{P}$. By (2), FL has the smallest expected loss among all permutation invariant strategies for any $\boldsymbol{P}$.

Theorem 6. *Let $\{P_\theta : \theta \in \Theta\}$ be a one-dimensional exponential family, and let $\boldsymbol{\omega}^{\mathrm{fl}} = (\boldsymbol{w}_1^{\mathrm{fl}}, \dots, \boldsymbol{w}_T^{\mathrm{fl}})$ be the FL strategy. Then, for any K-vector $(\theta_1, \dots, \theta_K) \in \Theta^K$, and associated K-vector of distributions $\boldsymbol{P} = (P_{\theta_1}, \dots, P_{\theta_K})$, for any strategy $\boldsymbol{\omega} = (\boldsymbol{w}_1, \dots, \boldsymbol{w}_T)$, and any $t = 1, \dots, T$:*

$$\overline{\mathrm{loss}}_t(\boldsymbol{w}_t^{\mathrm{fl}}, \boldsymbol{P}) \le \overline{\mathrm{loss}}_t(\boldsymbol{w}_t, \boldsymbol{P}).$$

In particular, by (2), for any permutation invariant strategy $\boldsymbol{\omega}$:

$$\mathbb{E}_{\boldsymbol{P}}\left[\boldsymbol{w}_t^{\mathrm{fl}}(\boldsymbol{\ell}^{t-1}) \cdot \boldsymbol{\ell}_t\right] \le \mathbb{E}_{\boldsymbol{P}}\left[\boldsymbol{w}_t(\boldsymbol{\ell}^{t-1}) \cdot \boldsymbol{\ell}_t\right].$$

The proof is similar to the proof of Theorem 3 and is given in Appendix B.

The consequence of Theorem 6 is the minimaxity of FL strategy under exponential family model. However, contrary to the case of binary losses in Section 3, the minimax regret, pseudo-regret and risk can be infinite when the losses are unbounded. For instance, take $K = 2$ experts and $T = 1$ trial, and consider one-dimensional Gaussian family model with unit variance given above. Let μ_1 and μ_2 be the (mean-value) parameters of expert 1 and expert 2. By symmetry, the best any algorithm can do in the first trial is to put equal weights on both experts, so that the expected loss of the algorithm in the first trial is $\frac{1}{2}\mathbb{E}[\ell_{1,1}] + \frac{1}{2}\mathbb{E}[\ell_{2,1}] = \frac{1}{2}(\mu_1 + \mu_2)$. Assuming (without loss of generality) $\mu_1 \le \mu_2$, this means that the risk and pseudo-regret are equal to $\frac{1}{2}(\mu_1 + \mu_2) - \mu_1 = \frac{1}{2}(\mu_2 - \mu_1)$, while the regret is at least that amount. Making the difference $\mu_2 - \mu_1$ as large as we want, it follows that the minimax risk, pseudo-regret and regret can all be made arbitrarily large in just a single trial.

To exclude such cases, we restrict the parameters chosen by the adversary to lie in a subset of parameter space $\Theta_0 \subseteq \Theta$, for which the minimax value is finite. For instance, in the Gaussian family model, we can choose $\Theta_0 = \{\theta : |\theta| \leq B\}$ for some $B > 0$, which would imply the boundedness of the distribution mean. While in principle we let the algorithm know Θ_0 in advance, the theorem below shows that the minimax strategy is the FL strategy, which does not make use of such knowledge at all.

Corollary 7. *Let $\mathcal{P} = \{P_\theta : \theta \in \Theta_0\}$ be a one-dimensional exponential family with its natural parameter space restricted to a non-empty subset of Θ, $\Theta_0 \subseteq \Theta$, for which the minimax value is finite:*

$$R^* := \inf_{\boldsymbol{\omega}} \sup_{\boldsymbol{P} \in \mathcal{P}^K} R(\boldsymbol{\omega}, \boldsymbol{P}) < \infty,$$

where R is the expected regret, pseudo-regret, or excess risk. Let $\boldsymbol{\omega}^{\text{fl}} = (\boldsymbol{w}_1^{\text{fl}}, \ldots, \boldsymbol{w}_T^{\text{fl}})$ be the FL strategy. Then, for any $(\theta_1, \ldots, \theta_K) \in \Theta_0^K$ and associated $\boldsymbol{P} = (P_{\theta_1}, \ldots, P_{\theta_K})$, and any permutation invariant strategy $\boldsymbol{\omega}$:

$$R(\boldsymbol{\omega}^{\text{fl}}, \boldsymbol{P}) \leq R(\boldsymbol{\omega}, \boldsymbol{P}),$$

which implies:

$$\sup_{\boldsymbol{P} \in \mathcal{P}^K} R(\boldsymbol{\omega}^{\text{fl}}, \boldsymbol{P}) = R^*.$$

Proof. The first statement follows from the second part of Theorem 6, and the fact that the "loss of the best expert" part on each side of inequality is the same.

The second statement immediately follows from the first statement and Theorem 2. □

5. Losses in a bounded interval

In this section, we consider the case $\mathcal{X} = [0, 1]$ of bounded loss vectors, and we let $\mathcal{P}$ be the set of *all* distributions on $[0, 1]$ (with some straightforward changes, our analysis actually applies to any bounded interval $[a, b]$, but we keep $[0, 1]$ range for the sake of simplicity). We give a modification of FL strategy and prove its minimaxity. We later justify the modification by arguing that the plain FL strategy is not minimax for this setup.

We remark that Corollary 7 from previous section, which states the minimaxity of FL for exponential family model, does not contradict the suboptimality of FL shown below. This is because the exponential family model *does not* cover the setup considered here, as the set of all distributions on $[0, 1]$ does not form an exponential family. Thus, in this section we are dealing with a more ambitious goal (although restricted to a bounded domain, as opposed to Section 4) of competing against an adversary, whose experts can play with *any* distribution over $\mathcal{X}$.

In what follows, we sometimes refer to losses in the range $[0, 1]$ as *continuous losses*, as opposed to *binary losses* from $\{0, 1\}$.

5.1. Binarized FL

The modification of FL is based on the procedure we call *binarization*. A similar trick has already been used in [23] to make Thompson Sampling work for non-binary rewards, and in [12] to deal with non-binary losses in a version of Follow the Perturbed Leader algorithm. We define a binarization of any loss value $\ell_{t,k} \in [0, 1]$ as a Bernoulli random variable $b_{t,k}$ which takes out value 1 with probability $\ell_{t,k}$ and value 0 with probability $1 - \ell_{t,k}$. In other words, we replace each continuous loss $\ell_{t,k}$ by a random binary outcome $b_{t,k}$, such that $\mathbb{E}[b_{t,k}] = \ell_{t,k}$. Note that if $\ell_{t,k} \in \{0, 1\}$, then $b_{t,k} = \ell_{t,k}$, i.e. binarization has no effect on losses which are already binary. Let us also define $\boldsymbol{b}_t = (b_{t,1}, \ldots, b_{t,K})$, where all K Bernoulli random variables $b_{t,k}$ are independent. Similarly, $\boldsymbol{b}^t$ denotes a binary loss sequence $\boldsymbol{b}_1, \ldots, \boldsymbol{b}_t$, where the binarization procedure is applied independently (with a fresh set of Bernoulli variables) for each trial t. Now, given the loss sequence $\boldsymbol{\ell}^{t-1}$, we define the *binarized FL* strategy $\boldsymbol{\omega}^{\text{bfl}}$ by:

$$\boldsymbol{w}_t^{\text{bfl}}(\boldsymbol{\ell}^{t-1}) = \mathbb{E}_{\boldsymbol{b}^{t-1}}\left[\boldsymbol{w}_t^{\text{fl}}(\boldsymbol{b}^{t-1})\right],$$

where $\boldsymbol{w}_t^{\text{fl}}(\boldsymbol{b}^{t-1})$ is the standard FL strategy applied to binarized losses $\boldsymbol{b}^{t-1}$, and the expectation is over internal randomization of the algorithm (binarization variables).[4]

[4] Instead of computing the expectation over binarization variables, one can simply draw these variables once and play with FL on such binarized losses. Due to linearity of the loss with respect to the weight vector, the expected performance of such strategy (with expectation with respect to internal randomization of the algorithm) is the same as that of binarized FL.

Note that if the set of distributions $\boldsymbol{P}$ has support only on $\{0, 1\}$, then $\boldsymbol{w}_t^{\mathrm{bfl}} \equiv \boldsymbol{w}_t^{\mathrm{fl}}$. On the other hand, these two strategies may differ significantly for non-binary losses. However, we will show that for any K-vector of distributions $\boldsymbol{P}$ (with support in $[0, 1]$), $\boldsymbol{w}_t^{\mathrm{bfl}}$ will behave in the same way as $\boldsymbol{w}_t^{\mathrm{fl}}$ would behave on some particular K-vector of distributions over binary losses. To this end, we introduce *binarization of a K-vector of distributions $\boldsymbol{P}$*, defined as $\boldsymbol{P}^{\mathrm{bin}} = (P_1^{\mathrm{bin}}, \ldots, P_K^{\mathrm{bin}})$, where P_k^{bin} is a distribution with support $\{0, 1\}$ such that:

$$\mathbb{E}_{P_k^{\mathrm{bin}}}[\ell_{t,k}] = P_k^{\mathrm{bin}}(\ell_{t,k} = 1) = \mathbb{E}_{P_k}[\ell_{t,k}].$$

In other words, P_k^{bin} is a Bernoulli distribution which has the same expectation as the original distribution (over continuous losses) P_k.

Lemma 8. *For any K-vector of distributions $\boldsymbol{P} = (P_1, \ldots, P_K)$ with support on $\mathcal{X} = [0, 1]$,*

$$\mathbb{E}_{\boldsymbol{\ell}^t \sim \boldsymbol{P}}\left[\boldsymbol{w}_t^{\mathrm{bfl}}(\boldsymbol{\ell}^{t-1}) \cdot \boldsymbol{\ell}_t\right] = \mathbb{E}_{\boldsymbol{\ell}^t \sim \boldsymbol{P}^{\mathrm{bin}}}\left[\boldsymbol{w}_t^{\mathrm{fl}}(\boldsymbol{\ell}^{t-1}) \cdot \boldsymbol{\ell}_t\right].$$

Proof. Let μ_k be the expectation of $\ell_{t,k}$ according to either P_k or P_k^{bin}, $\mu_k := \mathbb{E}_{P_k}[\ell_{t,k}] = \mathbb{E}_{P_k^{\mathrm{bin}}}[\ell_{t,k}]$. Since for any prediction strategy $\boldsymbol{\omega}$, $\boldsymbol{w}_t$ depends on $\boldsymbol{\ell}^{t-1}$ and does not depend on $\boldsymbol{\ell}_t$, we have:

$$\mathbb{E}_{\boldsymbol{P}}\left[\boldsymbol{w}_t^{\mathrm{bfl}} \cdot \boldsymbol{\ell}_t\right] = \mathbb{E}_{\boldsymbol{P}}\left[\boldsymbol{w}_t^{\mathrm{bfl}}\right] \cdot \mathbb{E}_{\boldsymbol{P}}\left[\boldsymbol{\ell}_t\right] = \mathbb{E}_{\boldsymbol{P}}\left[\boldsymbol{w}_t^{\mathrm{bfl}}\right] \cdot \boldsymbol{\mu},$$

where $\boldsymbol{\mu} = (p_1, \ldots, p_K)$. Similarly,

$$\mathbb{E}_{\boldsymbol{P}^{\mathrm{bin}}}\left[\boldsymbol{w}_t^{\mathrm{fl}} \cdot \boldsymbol{\ell}_t\right] = \mathbb{E}_{\boldsymbol{P}^{\mathrm{bin}}}\left[\boldsymbol{w}_t^{\mathrm{fl}}\right] \cdot \boldsymbol{\mu}.$$

Hence, we only need to show that $\mathbb{E}_{\boldsymbol{P}}\left[\boldsymbol{w}_t^{\mathrm{bfl}}\right] = \mathbb{E}_{\boldsymbol{P}^{\mathrm{bin}}}\left[\boldsymbol{w}_t^{\mathrm{fl}}\right]$. This holds because $\boldsymbol{w}_t^{\mathrm{bfl}}$ "sees" only the binary outcomes resulting from the joint distribution of $\boldsymbol{P}$ and the distribution of binarization variables:

$$\mathbb{E}_{\boldsymbol{\ell}^{t-1} \sim \boldsymbol{P}}\left[\boldsymbol{w}_t^{\mathrm{bfl}}(\boldsymbol{\ell}^{t-1})\right] = \mathbb{E}_{\boldsymbol{\ell}^{t-1} \sim \boldsymbol{P}, \boldsymbol{b}^{t-1}}\left[\boldsymbol{w}_t^{\mathrm{fl}}(\boldsymbol{b}^{t-1})\right],$$

and for any $b_{t,k}$, the probability (jointly over P_k and the binarization variables) of $b_{t,k} = 1$ is the same as probability of $\ell_{t,k} = 1$ over the distribution P_k^{bin}:

$$\begin{aligned} P(b_{t,k} = 1) &= \int_{[0,1]} P(b_{t,k} = 1 | \ell_{t,k}) P_k(\ell_{t,k}) \mathrm{d}\ell_{t,k} \\ &= \int_{[0,1]} \ell_{t,k} P_k(\ell_{t,k}) \mathrm{d}\ell_{t,k} \;=\; \mu_k = P^{\mathrm{bin}}(\ell_{t,k} = 1). \end{aligned} \tag{5}$$

Hence,

$$\mathbb{E}_{\boldsymbol{\ell}^{t-1} \sim \boldsymbol{P}, \boldsymbol{b}^{t-1}}\left[\boldsymbol{w}_t^{\mathrm{fl}}(\boldsymbol{b}^{t-1})\right] = \mathbb{E}_{\boldsymbol{\ell}^{t-1} \sim \boldsymbol{P}^{\mathrm{bin}}}\left[\boldsymbol{w}_t^{\mathrm{fl}}(\boldsymbol{\ell}^{t-1})\right]. \qquad \square$$

Lemma 9. *For any K-vector of distributions $\boldsymbol{P} = (P_1, \ldots, P_K)$ with support on $\mathcal{X} = [0, 1]$,*

$$R(\boldsymbol{\omega}^{\mathrm{bfl}}, \boldsymbol{P}) \leq R(\boldsymbol{\omega}^{\mathrm{fl}}, \boldsymbol{P}^{\mathrm{bin}}),$$

where R is either the expected regret, the pseudo-regret, or the excess risk (for pseudo-regret and excess risk, the lemma holds with equality).

Proof. Lemma 8 shows that the expected loss of $\boldsymbol{\omega}^{\mathrm{bfl}}$ on $\boldsymbol{P}$ is the same as the expected loss of $\boldsymbol{\omega}^{\mathrm{fl}}$ on $\boldsymbol{P}^{\mathrm{bin}}$. Hence, to prove the inequality, we only need to consider the "loss of the best expert" part of each measure. For the pseudo-regret, and the expected regret, it directly follows from the definition of $\boldsymbol{P}^{\mathrm{bin}}$ that for any t, k, $\mathbb{E}_{\boldsymbol{P}}[\ell_{t,k}] = \mathbb{E}_{\boldsymbol{P}^{\mathrm{bin}}}[\ell_{t,k}]$, hence $\min_k \mathbb{E}_{\boldsymbol{P}}[\ell_{T,k}] = \min_k \mathbb{E}_{\boldsymbol{P}^{\mathrm{bin}}}[\ell_{T,k}]$, and similarly, $\min_k \mathbb{E}_{\boldsymbol{P}}[L_{T,k}] = \min_k \mathbb{E}_{\boldsymbol{P}^{\mathrm{bin}}}[L_{T,k}]$. Thus, for the pseudo-regret and the excess risk, the lemma actually holds with equality.

For the expected regret, we will show that $\mathbb{E}_{\boldsymbol{P}}[\min_k L_{T,k}] \geq \mathbb{E}_{\boldsymbol{P}^{\mathrm{bin}}}[\min_k L_{T,k}]$, which will finish the proof. Denoting $B_{T,k} = \sum_{t=1}^{T} b_{t,k}$, we have:

$$
\begin{aligned}
\mathbb{E}_{\boldsymbol{\ell}^T \sim \boldsymbol{P}^{\mathrm{bin}}}[\min_k L_{T,k}] &= \mathbb{E}_{\boldsymbol{\ell}^T \sim \boldsymbol{P}, \boldsymbol{b}^T}[\min_k B_{T,k}] \\
&\leq \mathbb{E}_{\boldsymbol{\ell}^T \sim \boldsymbol{P}}\left[\min_k \mathbb{E}_{\boldsymbol{b}^T}[B_{T,k}|\boldsymbol{\ell}^T]\right] \\
&= \mathbb{E}_{\boldsymbol{\ell}^T \sim \boldsymbol{P}}[\min_k L_{T,k}],
\end{aligned}
$$

where the first equality is from the fact that for any $b_{t,k}$, the probability (jointly over P_k and the binarization variables) of $b_{t,k} = 1$ is the same as probability of $\ell_{t,k} = 1$ with respect to distribution P_k^{bin} (see (5) in the proof of Lemma 8), while the inequality follows from Jensen's inequality applied to the concave function $\min(\cdot)$. □

Theorem 10. *Let $\boldsymbol{\omega}^{\mathrm{bfl}} = (\boldsymbol{w}_1^{\mathrm{bfl}}, \ldots, \boldsymbol{w}_T^{\mathrm{bfl}})$ be the binarized FL strategy. Then:*

$$
\sup_{\boldsymbol{P}} R(\boldsymbol{\omega}^{\mathrm{bfl}}, \boldsymbol{P}) = \inf_{\boldsymbol{\omega}} \sup_{\boldsymbol{P}} R(\boldsymbol{\omega}, \boldsymbol{P}),
$$

where R is the expected regret, pseudo-regret, or excess risk, the supremum is over all K-vectors of distributions on $[0, 1]$, and the infimum is over all prediction strategies.

Proof. Lemma 9 states that for any K-vector of distributions $\boldsymbol{P}$, $R(\boldsymbol{\omega}^{\mathrm{bfl}}, \boldsymbol{P}) \leq R(\boldsymbol{\omega}^{\mathrm{fl}}, \boldsymbol{P}^{\mathrm{bin}})$. Furthermore, since $\boldsymbol{\omega}^{\mathrm{bfl}}$ is the same as $\boldsymbol{\omega}^{\mathrm{fl}}$ when all the losses are binary, $R(\boldsymbol{\omega}^{\mathrm{bfl}}, \boldsymbol{P}^{\mathrm{bin}}) = R(\boldsymbol{\omega}^{\mathrm{fl}}, \boldsymbol{P}^{\mathrm{bin}})$, and hence $R(\boldsymbol{\omega}^{\mathrm{bfl}}, \boldsymbol{P}) \leq R(\boldsymbol{\omega}^{\mathrm{bfl}}, \boldsymbol{P}^{\mathrm{bin}})$, i.e. for every $\boldsymbol{P}$ over continuous losses, there is a corresponding $\boldsymbol{P}^{\mathrm{bin}}$ over binary losses which incurs at least the same regret/pseudo-regret/risk to $\boldsymbol{\omega}^{\mathrm{bfl}}$. Therefore,

$$
\sup_{\boldsymbol{P} \text{ on } [0,1]} R(\boldsymbol{\omega}^{\mathrm{bfl}}, \boldsymbol{P}) = \sup_{\boldsymbol{P} \text{ on } \{0,1\}} R(\boldsymbol{\omega}^{\mathrm{bfl}}, \boldsymbol{P}) = \sup_{\boldsymbol{P} \text{ on } \{0,1\}} R(\boldsymbol{\omega}^{\mathrm{fl}}, \boldsymbol{P}).
$$

By the second part of Corollary 5, for any prediction strategy $\boldsymbol{\omega}$:

$$
\sup_{\boldsymbol{P} \text{ on } \{0,1\}} R(\boldsymbol{\omega}^{\mathrm{fl}}, \boldsymbol{P}) \leq \sup_{\boldsymbol{P} \text{ on } \{0,1\}} R(\boldsymbol{\omega}, \boldsymbol{P}) \leq \sup_{\boldsymbol{P} \text{ on } [0,1]} R(\boldsymbol{\omega}, \boldsymbol{P}),
$$

which finishes the proof. □

Theorem 10 states that the binarized FL strategy is the minimax prediction strategy when the losses are continuous on $[0, 1]$. Note that the same arguments would hold for any other loss range $[a, b]$, where the binarization on losses would convert continuous losses to the binary losses with values in $\{a, b\}$.

5.2. Vanilla FL is not minimax for continuous losses

We introduced the binarization procedure to show that the resulting binarized FL strategy is minimax for continuous losses. So far, however, we did not exclude the possibility that the plain FL strategy (without binarization) could also be minimax in the continuous setup. In this section, we prove (by giving a counterexample) that this is not the case, so that the binarization procedure is justified. We will only consider excess risk for simplicity, but one can use similar arguments to show a counterexample for the expected regret and the pseudo-regret as well.

The counterexample proceeds by choosing the simplest non-trivial setup of $K = 2$ experts and $T = 2$ trials. We will first consider the case of binary losses and determine the minimax excess risk. Take two distributions P_1, P_2 on binary losses and denote $\mu_1 = P_1(\ell_{t,1} = 1)$ and $\mu_2 = P_2(\ell_{t,2} = 1)$, assuming (without loss of generality) that $\mu_1 \leq \mu_2$. The excess risk of the FL strategy (its expected loss in the second trial minus the expected loss of the first expert) is given by:

$$
P(\ell_{1,1} = 0, \ell_{1,2} = 1)\mu_1 + P(\ell_{1,2} = 0, \ell_{1,1} = 1)\mu_2 + P(\ell_{1,1} = \ell_{1,2})\frac{\mu_1 + \mu_2}{2} - \mu_1,
$$

which can be rewritten as:

$$
\underbrace{\mu_2(1-\mu_1)\mu_1 + \mu_1(1-\mu_2)\mu_2}_{=2\mu_1\mu_2 - \mu_1\mu_2(\mu_1+\mu_2)} + \underbrace{\Big(\mu_1\mu_2 + (1-\mu_1)(1-\mu_2)\Big)\frac{\mu_1+\mu_2}{2}}_{=\mu_1\mu_2(\mu_1+\mu_2) - (\mu_1+\mu_2)^2 + \frac{\mu_1+\mu_2}{2}} - \mu_1 = \frac{\mu_2 - \mu_1}{2} - \frac{(\mu_2-\mu_1)^2}{2}.
$$

Denoting $\delta = \mu_2 - \mu_1$, the excess risk can be concisely written as $\frac{\delta}{2} - \frac{\delta^2}{2}$. Maximizing over δ gives $\delta^* = \frac{1}{2}$ and hence *the maximum risk of FL on binary losses is equal to* $\frac{1}{8}$.

Now, the crucial point to note is that *this is also the minimax risk on continuous losses*. This follows because the binarized FL strategy (which is the minimax strategy on continuous losses) achieves the maximum risk on binary losses (for which it is equivalent to the FL strategy), as follows from the proof of Theorem 10. What remains to be shown is that there exist distributions P_1, P_2 on continuous losses which force FL to suffer more excess risk than $\frac{1}{8}$. We take P_1 with support on two points $\{\epsilon, 1\}$, where ϵ is a very small positive number, and $p_1 = P_1(\ell_{t,1} = 1)$. Note that $\mu_1 := \mathbb{E}[\ell_{t,1}] = p_1 + \epsilon(1 - p_1)$. P_2 has support on $\{0, 1-\epsilon\}$, and let $p_2 = P_2(\ell_{t,2} = 1-\epsilon)$, which means that $\mu_2 := \mathbb{E}[\ell_{t,2}] = p_2(1-\epsilon)$. We also assume $\mu_1 < \mu_2$ i.e. expert 1 is the "better" expert, which translates to $p_1 + \epsilon(1-p_1) < p_2(1-\epsilon)$. The main idea in this counterexample is that by using ϵ values, all "ties" are resolved in favor of expert 2, which makes the FL algorithm suffer more loss. More precisely, this risk of FL is now given by:

$$p_2(1-p_1)p_1 + p_1(1-p_2)p_2 + \underbrace{\Big(p_1 p_2 + (1-p_1)(1-p_2)\Big)p_2}_{\text{ties}} - p_1 + O(\epsilon).$$

Choosing, e.g. $p_1 = 0$ and $p_2 = 0.5$, gives $\frac{1}{4} + O(\epsilon)$ excess risk, which is more than $\frac{1}{8}$, given that we take ϵ sufficiently small.

6. Dependent experts

All setups discussed so far concern distributions over loss vectors which are i.i.d. between trials, but also i.i.d. between experts. From the practical point of view, it is clearly interesting to look at the case in which the adversary can choose any joint distribution over loss vectors (still i.i.d. between trials, but not necessarily i.i.d. between experts). Unfortunately, this case seems to be notoriously hard to approach, even in the binary loss case, and all our methods based on permutation invariance fail. Furthermore, we managed to find a numerical counterexample which shows that for binary losses $\boldsymbol{\ell}_t \in \{0,1\}^K$, $K = 3$ experts and $T = 3$ trials, the FL strategy is not minimax in terms of the expected regret.[5] This suggests that, similarly to the case of Section 5, at least some modification to FL is required.

In this section we present the simplest variant of dependent expert case, where all loss vectors are *unit* in the sense that $\boldsymbol{\ell}_t = \boldsymbol{e}_k$ for some $k = 1, \dots, K$ (only a single expert gets loss at a time). In other words, losses follow a multinomial distribution over K outcomes, which is clearly not i.i.d. between experts. Since now we consider a single joint (multinomial) distributions over loss vectors, we let P denote such distribution, with $p_k := P(\boldsymbol{\ell}_t = \boldsymbol{e}_k)$. The set of all multinomial distributions over K outcomes is denoted by $\mathcal{P}$. As $\mathcal{P}$ is closed under taking permutations over experts, one can easily show that all properties of permutation-invariant strategies proved in Section 2.2 also hold in this setup.

We now show that in the multinomial setting, the FL strategy retains the minimax properties analogous to those given in Theorem 3 and Theorem 6.

Theorem 11. *Let $\mathcal{P}$ be the family of multinomial distributions, and let $\boldsymbol{\omega}^{\mathrm{fl}}$ be the FL strategy. Then, for any $P \in \mathcal{P}$, any strategy $\boldsymbol{\omega}$, and any $t = 1, \dots, T$:*

$$\overline{\mathrm{loss}}_t(\boldsymbol{w}_t^{\mathrm{fl}}, P) \le \overline{\mathrm{loss}}_t(\boldsymbol{w}_t, P).$$

In particular, by (2), for any permutation invariant strategy $\boldsymbol{\omega}$:

$$\mathbb{E}_P\left[\boldsymbol{w}_t^{\mathrm{fl}}(\boldsymbol{\ell}^{t-1}) \cdot \boldsymbol{\ell}_t\right] \le \mathbb{E}_P\left[\boldsymbol{w}_t(\boldsymbol{\ell}^{t-1}) \cdot \boldsymbol{\ell}_t\right].$$

The proof is given in Appendix C. As in the case of binary and exponential family models, the consequence is:

Corollary 12. *Let $\mathcal{P}$ be the family of multinomial distributions over K, and let $\boldsymbol{\omega}^{\mathrm{fl}}$ be the FL strategy. Then, for any $P \in \mathcal{P}$, and any permutation invariant strategy $\boldsymbol{\omega}$:*

$$R(\boldsymbol{\omega}^{\mathrm{fl}}, P) \le R(\boldsymbol{\omega}, P),$$

where R is the expected regret, pseudo-regret, or excess risk. This implies:

$$\sup_{P \in \mathcal{P}} R(\boldsymbol{\omega}^{\mathrm{fl}}, P) = \inf_{\boldsymbol{\omega}} \sup_{P \in \mathcal{P}} R(\boldsymbol{\omega}, P).$$

[5] The counterexample is based on numerical maximization of the expected regret over all distributions on 8 loss vectors $(0,0,0), (0,0,1), (0,1,0), \dots$, where we found out that a slight modification of FL strategy gives a better worst-case expected regret: in trial $t = 3$, after seeing two loss vectors $(0,0,1)$ and $(1,1,0)$ (or any expert-permutation thereof), put $2/3$ weight on expert 3 and $1/6$ weight on experts 1 and 2 (FL would put equal weight $1/3$ on all experts). This decreases worst-case expected regret from 0.91956 (FL regret) to 0.91667.

7. Conclusions and open problem

In this paper, we determined the minimax strategy for the stochastic setting of prediction with expert advice in which each expert generates its losses i.i.d. according to some distribution. Interestingly, the minimaxity is achieved by a single strategy, simultaneously for three considered performance measures: the expected regret, the pseudo-regret, and the excess risk. We showed that when the losses are binary, the Follow the Leader algorithm is the minimax strategy for this game, and furthermore, it also has the smallest expected regret, pseudo-regret, and excess risk among all permutation invariant prediction strategies for *every* distribution over the binary losses simultaneously, even among (permutation invariant) strategies which know the distributions of the losses. These minimax properties of FL also generalize to the case in which each expert generates its losses from a distribution belonging to a one-dimensional exponential family (of which the case of binary losses is a special case), as well as in a simple setup of dependent experts, where the loss vectors follow multinomial distribution, i.e. only a single expert gets loss in a given trial.

We also showed that the FL strategy becomes suboptimal in the case of losses in the range $[0,1]$, in which each expert generates losses from an arbitrary distribution over $[0,1]$. However, by applying "binarization trick" to the losses, and using FL on the binarized sequence (which results in the *binarized FL* strategy), we obtain the minimax strategy in this setup.

Open problems. Using the notion of permutation invariance, we proved the optimality of the FL strategy for losses generated from an exponential family (which includes binary losses) independently over experts, and for loss vectors generated jointly from a multinomial distribution. The common property in all these cases is that the past cumulative losses of experts form a *sufficient statistic* [13,14], i.e. the cumulative losses capture all the information about the expert distributions. We would like to know whether the minimaxity of FL ever holds beyond this setup.

This paper concerns distributions over loss vectors which are i.i.d. between trials and i.i.d. between experts (except for a simple multinomial model where only one expert gets loss at a time). It would be interesting to determine the minimax strategy in a more general setting, when the adversary can choose any joint distribution over loss vectors in $[0,1]^K$, or at least in $\{0,1\}^K$ (still i.i.d. between trials, but not necessarily i.i.d. between experts). While our computational experiment mentioned in Section 6 suggests that vanilla FL is not optimal in this setting, it remains open whether an appropriate modification of FL (analogous to the binarized FL strategy) could reduce the problem to a simpler problem of independent binary losses or multinomial loss vectors, and thus give the minimax strategy. Finally, even more ambitious open problem would be to find a generic minimax strategy (if exists) for the case of joint distributions over loss vectors from a general compact set $\mathcal{X} \subset \mathbb{R}^K$.

Appendix A. Proof of Lemma 4

Lemma 4. *For any real numbers $x_1, \ldots, x_K, y_1, \ldots, y_K$, and any monotonically increasing function f, the expression:*

$$h(k) = \sum_{\sigma \in S_K} e^{\sum_{j=1}^K x_j f(y_{\sigma(j)})} y_{\sigma(k)}$$

is minimized by any $k^* \in \operatorname{argmin}_{j=1,\ldots,K} x_j$.

Proof. We proceed by induction on K. Take $K=2$ and choose any $k^* \in \operatorname{argmin}_{j=1,2} x_j$. Without loss of generality assume $k^*=1$, so that $x_1 \leq x_2$. We need to show that $h(1) \leq h(2)$, which for $K=2$ amounts to show that:

$$e^{x_1 f(y_1)+x_2 f(x_2)} y_1 + e^{x_1 f(y_2)+x_2 f(y_1)} y_2 \leq e^{x_1 f(y_1)+x_2 f(y_2)} y_2 + e^{x_1 f(y_2)+x_2 f(y_1)} y_1,$$

or, after rearranging the terms, that:

$$\left(e^{x_1 f(y_1)+x_2 f(y_2)} - e^{x_1 f(y_2)+x_2 f(y_1)}\right)(y_1 - y_2) \leq 0. \tag{A.1}$$

If $y_1 = y_2$, (A.1) holds trivially. For $y_1 > y_2$, we have $f(y_1) \geq f(y_2)$ (due to monotonicity of f), and hence:

$$\Big(x_1 f(y_1) + x_2 f(y_2)\Big) - \Big(x_1 f(y_2) + x_2 f(y_1)\Big) = (x_1 - x_2)(f(y_1) - f(y_2)) \leq 0,$$

because $x_1 \leq x_2$. This implies $e^{x_1 f(y_1)+x_2 f(y_2)} \leq e^{x_1 f(y_2)+x_2 f(y_1)}$, and (A.1) follows. Finally, when $y_1 < y_2$, we have $f(y_1) \leq f(y_2)$, and by similar arguments $e^{x_1 f(y_1)+x_2 f(y_2)} \geq e^{x_1 f(y_2)+x_2 f(y_1)}$, which gives (A.1). This finishes the proof of the base case $K=2$.

Now, choose any $K \geq 3$. Take any $k^* \in \operatorname{argmin}_{j=1,\ldots,K} x_j$, and any other $k \neq k^*$. We will show that $h(k^*) \leq h(k)$ which suffices to prove the lemma. Without loss of generality assume that $k^* \neq K$ and $k \neq K$ (it is always possible to rearrange indices this way, as $K \geq 3$). We expand the sum over permutations in the definition of $h(k^*)$ with respect to the value of $\sigma(K)$:

$$h(k^*) = \sum_{s=1}^{K} \sum_{\sigma\,:\,\sigma(K)=s} e^{\sum_{j=1}^{K} x_j f(y_{\sigma(j)})} y_{\sigma(k^*)}$$

$$= \sum_{s=1}^{K} e^{x_K f(y_s)} \sum_{\sigma\,:\,\sigma(K)=s} e^{\sum_{j=1}^{K-1} x_j f(y_{\sigma(j)})} y_{\sigma(k^*)}. \tag{A.2}$$

Fix s and define $K-1$ real numbers $y'_1, \dots, y'_{K-1}$ by:

$$y'_1 = y_1, \quad \dots, \quad y'_{s-1} = y_{s-1}, \quad y'_s = y_{s+1}, \quad \dots, \quad y'_{K-1} = y_K.$$

Note that:

$$\sum_{\sigma\,:\,\sigma(K)=s} e^{\sum_{j=1}^{K-1} x_j f(y_{\sigma(j)})} y_{\sigma(k^*)} = \sum_{\sigma \in S_{K-1}} e^{\sum_{j=1}^{K-1} x_j f(y'_{\sigma(j)})} y'_{\sigma(k^*)}. \tag{A.3}$$

Now, by inductive assumption we invoke the lemma for $2(K-1)$ real numbers $x_1, \dots, x_{K-1}, y'_1, \dots, y'_{K-1}$ to conclude that:

$$\sum_{\sigma \in S_{K-1}} e^{\sum_{j=1}^{K-1} x_j f(y'_{\sigma(j)})} y'_{\sigma(k^*)} \le \sum_{\sigma \in S_{K-1}} e^{\sum_{j=1}^{K-1} x_j f(y'_{\sigma(j)})} y'_{\sigma(k)},$$

where we used the fact that $k, k^* \le K-1$, and that $k^* = \operatorname{argmin}_{j=1,\dots,K-1} x_j$. By (A.3), it implies that:

$$\sum_{\sigma\,:\,\sigma(K)=s} e^{\sum_{j=1}^{K-1} x_j f(y_{\sigma(j)})} y_{\sigma(k^*)} \le \sum_{\sigma\,:\,\sigma(K)=s} e^{\sum_{j=1}^{K-1} x_j f(y_{\sigma(j)})} y_{\sigma(k)}.$$

Since s was arbitrary, we bounded each term in the sum over s in (A.2), and thus $h(k^*) \le h(k)$, which finishes the proof. □

Appendix B. Proof of Theorem 6

Theorem 6. *Let $\{P_\theta : \theta \in \Theta\}$ be a one-dimensional exponential family, and let $\boldsymbol{\omega}^{\mathrm{fl}} = (\boldsymbol{w}_1^{\mathrm{fl}}, \dots, \boldsymbol{w}_T^{\mathrm{fl}})$ be the FL strategy. Then, for any K-vector $(\theta_1, \dots, \theta_K) \in \Theta^K$, and associated K-vector of distributions $\boldsymbol{P} = (P_{\theta_1}, \dots, P_{\theta_K})$, for any strategy $\boldsymbol{\omega} = (\boldsymbol{w}_1, \dots, \boldsymbol{w}_T)$, and any $t = 1, \dots, T$:*

$$\overline{\mathrm{loss}}_t(\boldsymbol{w}_t^{\mathrm{fl}}, \boldsymbol{P}) \le \overline{\mathrm{loss}}_t(\boldsymbol{w}_t, \boldsymbol{P}).$$

In particular, by (2), for any permutation invariant strategy $\boldsymbol{\omega}$:

$$\mathbb{E}_{\boldsymbol{P}}\left[\boldsymbol{w}_t^{\mathrm{fl}}(\boldsymbol{\ell}^{t-1}) \cdot \boldsymbol{\ell}_t\right] \le \mathbb{E}_{\boldsymbol{P}}\left[\boldsymbol{w}_t(\boldsymbol{\ell}^{t-1}) \cdot \boldsymbol{\ell}_t\right].$$

Proof. First note that the joint probability of observing loss sequence $\boldsymbol{\ell}^t$ is given by,

$$P(\boldsymbol{\ell}^t) = \prod_{k=1}^{K} \prod_{q=1}^{t} h(\ell_{k,q}) e^{\theta_k \ell_{k,q} - \psi(\theta_k)} = h(\boldsymbol{\ell}^t) e^{\sum_{k=1}^{K} \theta_k L_{t,k} - t\psi(\theta_k)},$$

where we abbreviate $h(\boldsymbol{\ell}^t) = \prod_{k=1}^{K} \prod_{q=1}^{t} h(\ell_{k,q})$. We have:

$$\overline{\mathrm{loss}}_t(\boldsymbol{w}_t, \boldsymbol{P}) = \frac{1}{K!} \sum_{\sigma} \mathbb{E}_{\sigma(\boldsymbol{P})}\left[\boldsymbol{w}_t(\boldsymbol{\ell}^{t-1}) \cdot \boldsymbol{\ell}_t\right]$$

$$= \frac{1}{K!} \sum_{\sigma} \mathbb{E}_{\sigma(\boldsymbol{P})}\left[\boldsymbol{w}_t(\boldsymbol{\ell}^{t-1})\right] \cdot \mathbb{E}_{\sigma(\boldsymbol{P})}[\boldsymbol{\ell}_t]$$

$$= \frac{1}{K!} \sum_{\sigma} \int h(\boldsymbol{\ell}^{t-1}) e^{\sum_k \theta_{\sigma(k)} L_{t-1,k} - (t-1)\psi(\theta_{\sigma(k)})} \left(\sum_{k=1}^{K} w_{t,k}(\boldsymbol{\ell}^{t-1}) \mu_{\sigma(k)} \right) \mathrm{d}\boldsymbol{\ell}^{t-1}$$

$$= \frac{1}{K!} \int h(\boldsymbol{\ell}^{t-1}) e^{-(t-1)\sum_k \psi(\theta_k)} \underbrace{\sum_{k=1}^{K} w_{t,k}(\boldsymbol{\ell}^{t-1}) \left(\sum_{\sigma} e^{\sum_k \theta_{\sigma(k)} L_{t-1,k}} \mu_{\sigma(k)} \right)}_{=:\, \overline{\mathrm{loss}}_t(\boldsymbol{w}_t, \boldsymbol{P} | \boldsymbol{\ell}^{t-1})} \mathrm{d}\boldsymbol{\ell}^{t-1}.$$

We now fix $\boldsymbol{\ell}^{t-1}$ and consider the term $\overline{\mathrm{loss}}_t(\boldsymbol{w}_t, \boldsymbol{P}|\boldsymbol{\ell}^{t-1})$. To finish the proof, it suffices to show that for any $\boldsymbol{P}$, and any $\boldsymbol{\ell}^{t-1}$, $\overline{\mathrm{loss}}_t(\boldsymbol{w}_t, \boldsymbol{P}|\boldsymbol{\ell}^{t-1})$ is minimized by setting $\boldsymbol{w}_t = \boldsymbol{e}_{k^*}$ for any $k^* \in \mathrm{argmin}_j L_{t-1,j}$. Indeed, this implies that $\overline{\mathrm{loss}}_t(\boldsymbol{w}_t, \boldsymbol{P}|\boldsymbol{\ell}^{t-1})$ is also minimized by the FL strategy $\boldsymbol{w}_t^{\mathrm{fl}}$, which distributes its mass uniformly over all leaders, and since $\boldsymbol{w}_t^{\mathrm{fl}}$ minimizes $\overline{\mathrm{loss}}_t(\boldsymbol{w}_t, \boldsymbol{P}|\boldsymbol{\ell}^{t-1})$ for all $\boldsymbol{\ell}^{t-1}$, it also minimizes $\overline{\mathrm{loss}}_t(\boldsymbol{w}_t, \boldsymbol{P})$. As $\overline{\mathrm{loss}}_t(\boldsymbol{w}_t, \boldsymbol{P}|\boldsymbol{\ell}^{t-1})$ is linear in $\boldsymbol{w}_t$, it suffices to only check $\boldsymbol{w}_t = \boldsymbol{e}_k$, $k = 1, \ldots, K$, i.e. to show that for any $k^* \in \mathrm{argmin}_j L_{t-1,j}$ and any $k = 1, \ldots, K$, $\overline{\mathrm{loss}}_t(\boldsymbol{e}_{k^*}, \boldsymbol{P}|\boldsymbol{\ell}^{t-1}) \leq \overline{\mathrm{loss}}_t(\boldsymbol{e}_k, \boldsymbol{P}|\boldsymbol{\ell}^{t-1})$. But since:

$$\overline{\mathrm{loss}}_t(\boldsymbol{e}_k, \boldsymbol{P}|\boldsymbol{\ell}^{t-1}) = \sum_{\sigma} e^{\sum_{j=1}^K \theta_{\sigma(j)} L_{t-1,j}} \mu_{\sigma(k)},$$

this immediately follows from Lemma 4 with $x_j = L_{t-1,j}$, $y_j = \mu_j$ and $f(\mu) = \theta(\mu)$, which is monotonically increasing. This finishes the proof. □

Appendix C. Proof of Theorem 11

Theorem 11. *Let $\mathcal{P}$ be the family of multinomial distributions, and let $\boldsymbol{\omega}^{\mathrm{fl}}$ be the FL strategy. Then, for any $P \in \mathcal{P}$, any strategy $\boldsymbol{\omega}$, and any $t = 1, \ldots, T$:*

$$\overline{\mathrm{loss}}_t(\boldsymbol{w}_t^{\mathrm{fl}}, P) \leq \overline{\mathrm{loss}}_t(\boldsymbol{w}_t, P).$$

In particular, by (2), for any permutation invariant strategy $\boldsymbol{\omega}$:

$$\mathbb{E}_P\left[\boldsymbol{w}_t^{\mathrm{fl}}(\boldsymbol{\ell}^{t-1}) \cdot \boldsymbol{\ell}_t\right] \leq \mathbb{E}_P\left[\boldsymbol{w}_t(\boldsymbol{\ell}^{t-1}) \cdot \boldsymbol{\ell}_t\right].$$

Proof. Denote $p_k := P(\boldsymbol{\ell}_t = \boldsymbol{e}_k)$, so that $\mathbb{E}_P[\ell_{t,k}] = p_k$. We have:

$$\begin{aligned}
\overline{\mathrm{loss}}_t(\boldsymbol{w}_t, P) &= \frac{1}{K!} \sum_{\sigma} \mathbb{E}_{\sigma(P)}\left[\boldsymbol{w}_t(\boldsymbol{\ell}^{t-1}) \cdot \boldsymbol{\ell}_t\right] \\
&= \frac{1}{K!} \sum_{\sigma} \mathbb{E}_{\sigma(P)}\left[\boldsymbol{w}_t(\boldsymbol{\ell}^{t-1})\right] \cdot \mathbb{E}_{\sigma(P)}\left[\boldsymbol{\ell}_t\right] \\
&= \frac{1}{K!} \sum_{\sigma} \sum_{\boldsymbol{\ell}^{t-1}} \left(\prod_{k=1}^K p_{\sigma(k)}^{L_{t-1,k}}\right) \left(\sum_{k=1}^K w_{t,k}(\boldsymbol{\ell}^{t-1}) p_{\sigma(k)}\right) \\
&= \frac{1}{K!} \sum_{\boldsymbol{\ell}^{t-1}} \underbrace{\sum_{k=1}^K w_{t,k}(\boldsymbol{\ell}^{t-1}) \left(\sum_{\sigma} \prod_{j=1}^K p_{\sigma(j)}^{L_{t-1,j}} p_{\sigma(k)}\right)}_{=: \overline{\mathrm{loss}}_t(\boldsymbol{w}_t, P|\boldsymbol{\ell}^{t-1})}.
\end{aligned}$$

As in the proof of Theorem 3 and Theorem 6, it suffices to show that $\overline{\mathrm{loss}}_t(\boldsymbol{e}_k, P|\boldsymbol{\ell}^{t-1}) = \sum_\sigma \prod_{j=1}^K p_{\sigma(j)}^{L_{t-1,j}} p_{\sigma(k)}$ is minimized for $k^* = \mathrm{argmin}_j L_{t-1,j}$. If $p_j \neq 0$ for all j, this can be done by transforming:

$$\overline{\mathrm{loss}}_t(\boldsymbol{e}_k, P|\boldsymbol{\ell}^{t-1}) = \sum_{\sigma} e^{\sum_{j=1}^K L_{t-1,j} \log p_{\sigma(j)}} p_{\sigma(k)},$$

and invoking Lemma 4 with $x_j = L_{t-1,j}$, $y_j = p_j$ and $f(x) = \log x$, which is monotonically increasing. Since the theorem holds for $p_j = \epsilon$ with arbitrarily small ϵ, it must also hold for $p_j = 0$ by taking the limit $\epsilon \to 0$ and using the continuity of $\overline{\mathrm{loss}}_t(\boldsymbol{w}_t, P|\boldsymbol{\ell}^{t-1})$. □

References

[1] W. Kotłowski, On minimaxity of follow the leader strategy in the stochastic setting, in: Proceeding of the 27th International Conference on Algorithmic Learning Theory (ALT '16), in: Lecture Notes in Artificial Intelligence, vol. 9925, Springer-Verlag, 2016, pp. 261–275.
[2] N. Cesa-Bianchi, Y. Freund, D. Haussler, D.P. Helmbold, R.E. Schapire, M.K. Warmuth, How to use expert advice, J. ACM 44 (3) (1997) 427–485.
[3] N. Cesa-Bianchi, G. Lugosi, Prediction, Learning, and Games, Cambridge University Press, 2006.
[4] V. Vovk, A game of prediction with expert advice, J. Comput. System Sci. 56 (2) (1998) 153–173.
[5] N. Littlestone, M.K. Warmuth, The Weighted Majority algorithm, Inform. and Comput. 108 (2) (1994) 212–261.
[6] Y. Freund, R.E. Schapire, A decision-theoretic generalization of on-line learning and an application to boosting, J. Comput. System Sci. 55 (1997) 119–139.
[7] A. Kalai, S. Vempala, Efficient algorithms for online decision problems, J. Comput. System Sci. 71 (3) (2005) 291–307.
[8] J. Abernethy, M.K. Warmuth, J. Yellin, When random play is optimal against an adversary, in: COLT, 2008, pp. 437–445.

[9] W.M. Koolen, Combining Strategies Efficiently: High-Quality Decisions from Conflicting Advice, Ph.D. thesis, ILLC, University of Amsterdam, 2011.
[10] J. Abernethy, A. Agarwal, P.L. Bartlett, A. Rakhlin, A stochastic view of optimal regret through minimax duality, in: COLT, 2009.
[11] P.D. Grünwald, The Minimum Description Length Principle, MIT Press, Cambridge, MA, 2007.
[12] T. van Erven, W. Kotłowski, M.K. Warmuth, Follow the leader with dropout perturbations, in: COLT, 2014, pp. 949–974.
[13] T. Ferguson, Mathematical Statistics: A Decision Theoretic Approach, Academic Press, 1967.
[14] J.O. Berger, Statistical Decision Theory and Bayesian Analysis, Springer, 1985.
[15] S. de Rooij, T. van Erven, P.D. Grünwald, W.M. Koolen, Follow the leader if you can, hedge if you must, J. Mach. Learn. Res. 15 (1) (2014) 1281–1316.
[16] A. Sani, G. Neu, A. Lazaric, Exploiting easy data in online optimization, in: NIPS, 2014, pp. 810–818.
[17] W.M. Koolen, T. van Erven, Second-order quantile methods for experts and combinatorial games, in: COLT, 2015, pp. 1155–1175.
[18] H. Luo, R.E. Schapire, Achieving all with no parameters: AdaNormalHedge, in: COLT, 2015, pp. 1286–1304.
[19] R. Huang, T. Lattimore, A. György, C. Szepesvari, Following the leader and fast rates in linear prediction: curved constraint sets and other regularities, in: Advances in Neural Information Processing Systems, vol. 29, Curran Associates, Inc., 2016, pp. 4970–4978.
[20] S. Bubeck, N. Cesa-Bianchi, Regret analysis of stochastic and nonstochastic multi-armed bandit problems, Found. Trends Mach. Learn. 5 (1) (2012) 1–122.
[21] L. Devroye, L. Györfi, G. Lugosi, A Probabilistic Theory of Pattern Recognition, 1st edition, Springer-Verlag, 1996.
[22] O. Barndorff-Nielsen, Information and Exponential Families in Statistical Theory, Wiley, Chichester, UK, 1978.
[23] S. Agrawal, N. Goyal, Analysis of Thompson sampling for the multi-armed bandit problem, in: S. Mannor, N. Srebro, R.C. Williamson (Eds.), Proceedings of the 25th Annual Conference on Learning Theory, in: JMLR: Workshop and Conference Proceedings, vol. 23, 2012, pp. 39.1–39.26.