Fast rates for prediction with limited expert advice

El Mehdi Saad, Gilles Blanchard

Fast rates for prediction with limited expert advice

El Mehdi Saad[1], **Gilles Blanchard**[1,2]
[1]Laboratoire de Mathématiques d'Orsay, CNRS, Université Paris-Saclay; [2]Inria

Abstract

We investigate the problem of minimizing the excess generalization error with respect to the best expert prediction in a finite family in the stochastic setting, under limited access to information. We assume that the learner only has access to a limited number of expert advices per training round, as well as for prediction. Assuming that the loss function is Lipschitz and strongly convex, we show that if we are allowed to see the advice of only one expert per round for T rounds in the training phase, or to use the advice of only one expert for prediction in the test phase, the worst-case excess risk is $\Omega(1/\sqrt{T})$ with probability lower bounded by a constant. However, if we are allowed to see at least two actively chosen expert advices per training round and use at least two experts for prediction, the fast rate $\mathcal{O}(1/T)$ can be achieved. We design novel algorithms achieving this rate in this setting, and in the setting where the learner has a budget constraint on the total number of observed expert advices, and give precise instance-dependent bounds on the number of training rounds and queries needed to achieve a given generalization error precision.

Keywords: Online Learning, Budgeted Learning, Prediction with expert advice.

1 Introduction and setting

We consider a generic prediction problem in a stochastic setting: a target random variable Y taking values in $\mathcal{Y}$ is to be predicted by a user-determined forecast F, also modeled as a random variable, taking values in a closed convex subset $\mathcal{X}$ of $\mathbb{R}^d$. The mismatch between the two is measured via a loss function $l(F, Y)$. The quality of the agent's output is measured by its generalization risk

$$R(F) := \mathbb{E}\big[l(F, Y)\big].$$

To assist us in this task, the forecast or "advice" of a number of "experts" $(F_1, \dots, F_K)$ (also modeled as random variables) can be requested. The agent's objective is to achieve a risk as close as possible to the risk of the best expert $R^* = \min_{i\in[\![K]\!]} R(F_i)$ (for a nonnegative integer n, we denote $[\![n]\!] = \{1, \dots, n\}$). We measure the performance of the user's forecast via its excess risk (or average regret) with respect to that best expert.

The literature on expert advice generally considers the *cumulative* regret over a sequence of forecasts F_t followed by observation of the target variable Y_t and incurring the loss $l(F_t, Y_t)$, $t = 1, \dots, T$. In the present work we will separate observation (or training) phase and forecast phase: the user is allowed to observe (some of) the expert's predictions and the target variable for a number of independent, identically distributed rounds $(Y_t, F_{1,t}, \dots, F_{K,t})_{1\le t\le T}$ following certain rules to be specified. After the observation phase, the user must decide of a prediction strategy, namely a convex combination of the experts $\widehat{F} = \sum_{i=1}^k \widehat{w}_i F_i$, where the weights $\widehat{w}_i$ can be chosen based on the information gathered in the training phase. The risk of this strategy is $R(\widehat{F})$, where the risk is evaluated on new, independent data. In other words, if the training phase takes place over T independent rounds, the forecast risk is the expected loss over the $(T + 1)$th, independent, round.

35th Conference on Neural Information Processing Systems (NeurIPS 2021).

In some situations, it may be overly expensive to query the advice of all experts at each round. The cost can be monetary if each expert demands to be paid to reveal his opinion, possibly because they have access to some information that others do not. In this case we may have a total limit on how much we can spend. In a different context, it is unrealistic to ask for the advice of all available doctors or to run a large battery of tests on each patient. In this case, we may be have a strong limit on the number of expert opinions that can be consulted for each training instance. In a more typical machine learning scenario, each "expert" might be a fixed prediction method $F_i = f_i(X)$ (using the information of a covariate X), where the predictor functions f_i have been already trained in advance, albeit based on different sets of parameters or methodology; the goal then amounts to predictor selection or aggregation, in a situation where the computation of each single prediction constitutes the bottleneck cost, rather than data acquisition. Overall the agent's goal is to achieve a risk close to optimal while sparing on the number of experts queries – both at training time and for forecast.

Motivated by these questions we investigate several scenarios for prediction with limited access to expert advice. Furthermore, our emphasis is on obtaining *fast convergence rates* guarantees on the excess risk (i.e. $O(1/T)$ or $O(1/C)$, where C is the total query budget). These are possible under a strong convexity assumption of the loss, specified below. Our contributions are the following.

- As a preliminary, we revisit (Section 3) the *full information setting*, with no limitations on queries. Maybe surprisingly, we contribute a new algorithm that is both simpler than existing ones and for which the proof of the fast convergence rate for excess risk is also elementary. Furthermore, for forecast we only need to consult 2 experts. The general principle of this algorithm will be reused in the limited observation settings.
- We then investigate (Section 4) the *budgeted setting* where we have a total query budget constraint C for the training phase; then (Section 5) the *two-query* setting where the agent is limited to $m = 2$ queries per training round. In both cases, we give precise efficiency guarantees on the number of training expert queries needed to achieve a given precision for forecast. The obtained bounds come both in *instance-independent* (agnostic) and *instance-dependent* (depending on the experts' structure) flavors.
- Finally, we give some lower bounds (Section 6) were we show that fast rates cannot be achieved if the agent is only allowed to consult one single expert per training round *or* for forecast.

The following assumption on the loss will be made throughout the paper:

Assumption 1. *$\forall y \in \mathcal{Y}$: $x \in \mathcal{X} \subseteq \mathbb{R}^d \mapsto l(x, y)$ is L-Lipschitz and ρ-strongly convex.*

Recall that a function $f : \mathcal{X} \to \mathbb{R}$ is L-Lipschitz if $\forall x, y \in \mathcal{X}$:$|f(x) - f(y)| \leq L\|x - y\|$, and ρ-strongly convex if the function: $x \to f(x) - \frac{\rho^2}{2}\|x\|^2$ is convex.

Remarks. Assumption 1 implies that the diameter of $\mathcal{X}$ is bounded by $8L/\rho^2$ and the quantity $\sup_{x,x' \in \mathcal{X}, y \in \mathcal{Y}} |l(x, y) - l(x', y)|$ is bounded by $B := 8L^2/\rho^2$ (this notation shorthand will be used throughout the paper). Consequently, without loss of generality we can assume that the loss is bounded by B (see Lemma[1] S-1 and subsequent discussion for details). It is satisfied, for example, in the following setting: least square loss $l(x, y) = (y - x)^2$ where $x \in \mathcal{X}$ and $y \in \mathcal{Y}$ with $\mathcal{X}$ and $\mathcal{Y}$ are bounded subsets of $\mathbb{R}^d$. Prior knowledge on ρ is not necessary if L and an upper bound on the the l_∞ norm of the target variable Y and the experts are known.

2 Discussion of related Work

Games with limited feedback (slow rates): Our work investigates what happens between the full information and single-point feedback games. Learning with a restricted access to information was considered under various settings in [6], [19], [12], [20], [5]. A setting close to ours was considered in [21], where the agent chooses in each round a subset of experts to observe their advice, then follows the prediction of one expert. To minimize the cumulative regret in the adversarial setting, they used an extension of the Exp3 algorithm, which allows to have an excess risk of $\mathcal{O}(\sqrt{1/T})$ in the limited feedback setting and $\mathcal{O}(\sqrt{\log(C)/C})$ in the budgeted case with a budget C.

[1] References starting with a prefix S- point to the supplemental material.

The differences in the setting considered here is that (a) we are interested in the generalization error in the stochastic setting rather than the cumulative regret in an adversarial setting and (b) our assumptions of the convexity of the loss allow for the possibility of fast excess risk convergence. Moreover, we consider the more general case where the player is allowed to combine p out of K experts for prediction. The possibility of playing a subset of arms was considered in the literature of Multiple Play Multi-armed bandits. It was treated with a budget constraint in [26] for example (see also [24]), where at each round, exactly p out of K possible arms have to be played. In addition to observing the individual rewards for each arm played, the player also learns a vector of costs which has to be covered with an a-priori defined budget C. In the stochastic setting, a UCB-type procedure gives a bound for the cumulative regret of $\mathcal{O}(\Delta_{\min}^{-1}\log(C)/C)$ that holds only in expectation, where $\Delta_{\min}^{-1}$ denotes the gap between the best choice of arms and the second best choice. This bound leads to an instance dependent bound of $\mathcal{O}(\sqrt{\log(C)/C})$ in the worst case. In the adversarial setting, an extension of Exp3 procedure gives a bound of $\mathcal{O}(\sqrt{\log(C)/C})$ for the cumulative regret that holds with high probability. In another online problem, where the objective is to minimize the cumulative regret in an adversarial setting with a small effective range of losses, [11] have shown the impossibility of regret scaling with the effective range of losses in the bandit setting, while [23] showed that it is possible to circumvent this impossibility result if the player is allowed one additional observation per round. However, in the settings considered, it is impossible to achieve a regret dependence on T better than the rate of $\mathcal{O}(1/\sqrt{T})$.

Fast rates in the full information setting: The learning task of doing as well as the best expert of a finite family in the sense of generalization error has been studied quite extensively in the full information case. In an adversarial setting, it is well-known that under suitable assumptions on the loss function (typically related to strong convexity), an appropriately tuned exponential weighted average (EWA) strategy has cumulative regret bounded by the "fast rate" $\mathcal{O}(\log(K)/T)$ [13, 9, 4], which, combined with the online-to-batch conversion principle [8, 4] (also known as progressive mixture rule, [7, 25]), yields a bound of the same order for the *expected* excess prediction risk in the stochastic case. However, it was shown that progressive mixture type rules are *deviation suboptimal* for prediction [2], that is, their excess risk takes a value larger than $c/\sqrt{T}$ with constant positive probability over the training phase. To lift the apparent contradiction between the two last statements, consider that the excess risk of the EWA can take *negative* values, since it is an *improper* learning rule. Thus negative and positive "large" deviations can compensate each other so that the expectation is small. The inefficiency of EWA in deviation is a significant drawback, and alternatives to the EWA progressive mixture rule that achieve $\mathcal{O}(\log(K)/T)$ excess prediction risk with high probability were proposed by [17] and [3]. In [17], the strategy consists in whittling down the set of experts by elimination of obviously suboptimal experts, and performing empirical risk minimization (ERM) over the convex combinations of the remaining experts. In [3], the *empirical star* algorithm consists in performing an ERM over all segments consisting of a two-point convex combination of the ERM expert and any other expert. Note that the empirical star algorithm has the advantage that the final prediction rule is a convex combination of (at most) *two* experts.

Linear regression with partially observed attributes: Other related work is that of [10], and [14] on learning linear regression models with partially observed attributes. The most related setting to ours is the local budget setting, where the learner is allowed to output a linear combination of features for prediction. The key idea is to use the observed attributes in order to build an unbiased estimate of the full information sample, then to use an optimization procedure to minimize the penalized empirical loss. In our setting, the minimization of penalized empirical loss was shown to be suboptimal (see [16]). Moreover, while we want to predict as well as the best expert, in [10], the objective is to be as good as the best linear combination of features with a small additive term (the optimal rate, in this case, is $\mathcal{O}\big(1/\sqrt{T}\big)$). Finally, we consider that the restriction on observed attributes (experts advice) does not apply only to the training samples but also to the testing data.

Online convex optimization with limited feedback: The idea of using multiple point feedback to achieve faster rates appeared in the online convex optimization literature (see [1], and [22]). It was shown that in the setting where the adversary chooses a loss function in each round if the player is allowed to query this function in two points, it is possible to achieve minimax rates that are close to those achievable in the full information setting. The key idea is to build a randomized estimate of the gradients, which are then fed into standard first-order algorithms. These ideas are not convertible into our setting because we consider a non-convex set of experts.

3 The full information case

In this section, we revisit the "classical" case where there is no constraint on the number of expert queries per observation round; assume the output of all experts are observed for T rounds (in other words, T i.i.d. training examples), which is the full information or "batch" setting. We want to output a final prediction rule with prediction risk controlled with high probability over the training phase.

We start with putting forward an apparently new rule , simpler than existing ones [17, 3], for the full information setting which, like the empirical star [3], outputs a convex combination of two experts. In contrast to the latter, our rule does not need any optimization over a union of segments. The underlying principle will guide us to construct a budget efficient expert selection rule in the sequel.

Define $\hat{R}(F_i) := T^{-1}\sum_{t=1}^{T} l(F_{i,t}, Y_t)$ the empirical loss of expert i, and $\hat{d}_{ij} := (T^{-1}\sum_{t=1}^{T}(F_{i,t} - F_{j,t})^2)^{\frac{1}{2}}$ the empirical L_2 distance between experts i and j over T rounds. Finally let $\alpha = \alpha(\delta) := (\log(4K\delta^{-1})/T)^{\frac{1}{2}}$, where $\delta \in (0,1)$ is a fixed confidence parameter. Define

$$\Delta_{ij} := \hat{R}(F_j) - \hat{R}(F_i) - 6\alpha \max\left\{L\hat{d}_{ij}, B\alpha\right\}. \tag{1}$$

The quantity Δ_{ij} can be interpreted as a test statistic: if $\Delta_{ij} > 0$, then we have a guarantee that $R(F_j) > R(F_i)$, so that expert j is sub-optimal; this guarantee holds for all (i,j) uniformly with probability $(1-\delta)$. It therefore makes sense to reduce the set of candidates to

$$S := \Big\{ j \in [\![K]\!] : \sup_{j \in [\![K]\!]} \Delta_{ij} \leq 0 \Big\}. \tag{2}$$

Our new full information setting rule is the following:

$$\text{choose } \bar{k} \in S \text{ arbitrarily} ; \quad \text{pick } \bar{j} \in \operatorname*{Arg\,Max}_{j \in S} \hat{d}_{\bar{k}j}; \quad \text{predict } \widehat{F} := \frac{1}{2}(F_{\bar{k}} + F_{\bar{j}}). \tag{3}$$

In words, the above rule consists in eliminating all experts that are manifestly outperformed by another one, and, among the remaining experts, pick two that disagree as much as possible (in terms of empirical L^2 distance) and output their simple average for prediction. The next theorem establishes fast convergence rate for the excess risk of this rule:

Theorem 3.1. *If Assumption 1 holds and $\delta \in (0,1)$ is fixed, then for the prediction rule $\widehat{F}$ defined by* (3)*, it holds with probability $1-3\delta$ over the training phase (c is an absolute constant):*

$$R(\widehat{F}) \leq R^* + cB\frac{\log(4K\delta^{-1})}{T}.$$

Proof. Let $d_{ij}^2 = \mathbb{E}\big[(F_i - F_j)^2\big]$. The result hinges on the following high confidence control of risk differences, established in Corollary S-4 as a direct consequence of the empirical Bernstein's inequality: with probability at least $1-3\delta$, it holds:

$$\text{For all } i,j \in [\![K]\!] : \qquad \Delta_{ij} \leq (R_j - R_i) \leq \Delta_{ij} + 32\alpha\max(Ld_{ij}, B\alpha). \tag{4}$$

Let $i^* \in \operatorname{Arg\,Min}_{i \in [\![K]\!]} R_i$ be an optimal expert. Since $R_{i^*} - R_j \leq 0$ for all $j \in [\![K]\!]$, it follows that if (4) holds, then $i^* \in S$, from the definition of S. So if (4) holds, we have

$$\begin{aligned}
R\left(\frac{F_{\bar{k}} + F_{\bar{j}}}{2}\right) &\leq \frac{1}{2}\big(R_{\bar{k}} + R_{\bar{j}}\big) - \frac{\rho^2}{8}d_{\bar{k}\bar{j}}^2 \\
&= R^* + \frac{1}{2}\big((R_{\bar{k}} - R_{i^*}) + (R_{\bar{j}} - R_{i^*})\big) - \frac{\rho^2}{8}d_{\bar{k}\bar{j}}^2 \\
&\leq R^* + \frac{1}{2}\big(\Delta_{\bar{k}i^*} + \Delta_{\bar{j}i^*}\big) + 16\alpha\big(\max\big(Ld_{\bar{j}i^*}, B\alpha\big) + \max(Ld_{\bar{k}i^*}, B\alpha)\big) - \frac{\rho^2}{8}d_{\bar{k}\bar{j}}^2 \\
&\leq R^* + 32B\alpha^2 + 48L\alpha d_{\bar{k}\bar{j}} - \frac{\rho^2}{8}d_{\bar{k}\bar{j}}^2;
\end{aligned}$$

where we have used strong convexity of the loss (and therefore of $R(.)$ with respect to the L^2 distance) in the first line; the right-hand side of (4) in the third line; and, in the last line, the fact that $\bar{j}, \bar{k}, i^*$ are all in S along with $d_{\bar{j}i^*} \leq d_{\bar{j}\bar{k}} + d_{\bar{k}i^*} \leq 2d_{\bar{j}\bar{k}}$ by construction of $\bar{j}$. Finally upper bounding the value of the last bound by its maximum possible value as a function of $d_{\bar{k}\bar{j}}$ and recalling $B = 8L^2/\rho^2$, we obtain the statement. □

4 Budgeted Setting

In this section, we consider the budgeted setting. More precisely, given an a-priori defined budget C, at each round the decision-maker selects an arbitrary subset of experts and asks for their predictions. The choice of these experts may of course depend on past observations available to the agent. The player then pays a unit for each observed expert's advice. The game finishes when the budget is exhausted, at which point the player outputs a convex combination of experts for prediction.

We convert the batch rule defined in the full information setting to an "online" rule by performing the test $\Delta_{ji} > 0$ for each pair (i, j) after each allocation. If at any round an expert $i \in [\![K]\!]$ fails any of these tests (i.e $\exists j : \Delta_{ji} > 0$), it is no longer queried. This extension allows us to derive instance dependent bounds, which cover the rates obtained in the batch setting in the worst case.

Since the tests $\Delta_{ij} > 0$ are performed after each allocation, we introduce the following modification on the definition of Δ_{ij}, for concentration inequalities to hold uniformly over the runtime of the procedure. We define $\Delta_{ij}(t, \delta)$ as follows:

$$\Delta_{ij}(t,\delta) := \hat{R}(j,t) - \hat{R}(i,t) - 6\alpha(t, \delta/(t(t+1)) \max\left\{L\hat{d}_{ij}(t), B\alpha(t, \delta/(t(t+1))\right\}.$$

Algorithm 1 Budgeted aggregation

Input δ, L and ρ.
Initialization: $S \leftarrow [\![K]\!]$.
for $T = 1, 2, \dots$ **do**
  Jointly query all the experts in S and update $\Delta_{ij} > 0$ for all i, j.
  For all $i, j \in [\![K]\!]$, if $\Delta_{ij} > 0$, eliminate j: $S \leftarrow S \setminus \{j\}$.
  if the budget is consumed **then**
    let $\bar{k} \in S$, and $\bar{l} \leftarrow \underset{j \in S}{\text{argmax}}\ \hat{d}_{\bar{k}j}$.
    Return $\frac{1}{2}(F_{\bar{k}} + F_{\bar{l}})$.
  end if
end for

Let $\mathcal{S}^* := \text{Arg Min}_{i \in [\![K]\!]} R(F_i)$ denote the set of optimal experts. For $i, j \in [\![K]\!]$, we denote by $d_{ij} := (\mathbb{E}[(F_i - F_j)^2])^{1/2}$ the L_2 distance between the experts F_i and F_j. For $i \in [\![K]\!]$, we introduce the following quantity:

$$\Lambda_i := \min_{i^* \in \mathcal{S}^*} \max\left\{ \frac{L^2 d_{ii^*}^2}{|R(F_i) - R(F_{i^*})|^2}; \frac{B}{R(F_i) - R(F_{i^*})} \right\}.$$

Define the following set of experts:

$$\mathcal{S}_\epsilon = \left\{ i \in [\![K]\!] : \Lambda_i > \frac{1}{\epsilon} \right\},$$

and let $\mathcal{S}_\epsilon^c$ be its complementary.

Theorem 4.1. *(Instance dependent bound) Suppose Assumption 1 holds. Let $C \geq K$ denote the global budget on queries and denote $\hat{g}$ the output of Algorithm 1 with inputs (δ, L, ρ) when the budget C runs out. For any $\epsilon \geq 0$, if:*

$$C > 578 C_\epsilon \log\big(K\delta^{-1} C_\epsilon\big),$$

where

$$C_\epsilon := \sum_{i \in \mathcal{S}_\epsilon^c} \Lambda_i + |\mathcal{S}_\epsilon| \ \min\left\{ \frac{1}{\epsilon}; \Lambda^* \right\},$$

where $\Lambda^* := \max_{i:\Lambda_i < +\infty} \Lambda_i$, then, with probability at least $1 - \delta$:

$$R(\hat{g}) \leq R^* + cB\epsilon,$$

where c is an absolute constant.

Remarks. *Observe that the above result gives in particular a query budget bound for the problem of best expert identification in our setting, by taking $\epsilon = 0$, in which case the required expert query budget is of order $\sum_{i:\Lambda_i<+\infty} \Lambda_i$ up to logarithmic terms. We can compare this to the problem of best arm identification in a bandit setting (one arm pull/query per round); our setting can be cast into that framework by considering each expert as an arm and only recording the information of the loss of the asked expert. The known optimal query bound for best arm identification in the classical multi-armed bandits setting with loss/reward bounded by B is of order $\sum_{i:\Lambda_i<+\infty} \widetilde{\Lambda}_i$ [15], where $\widetilde{\Lambda}_i = B^2(R(F_i) - R(F_{i^*}))^{-2}$. Since the diameter of $\mathcal{X}$ is bounded by B/L (see Lemma S-1), it holds $\Lambda_i \leq \widetilde{\Lambda}_i$. Hence, for best expert identification, the bound of Theorem 4.1 improves upon the best arm identification bound, potentially by a significant margin (in particular concerning the contribution of suboptimal but close to optimal experts for which $d_{ii^*} \ll B/L$ and $R_i - R_{i^*} \ll B$). Again, the improvement is due to the Assumption 1 on the loss and the possibility to query several experts per round, which are not used when casting the problem as a classical bandit setting.*

5 Two queries per round ($m = p = 2$)

In this section, we suppose that the decision-maker is constrained to see only two experts' advice per round ($m = 2$). We suppose that the horizon is unknown; when the game is halted, the player outputs a convex combination of at most two experts ($p = 2$). We will show that the rates obtained are as good as in the full information case in its dependence on the number of rounds T.

Algorithm 2 works as follows. To circumvent the limitation of observing only two experts per round, in each round, we sample a pair $(i, j) \in S \times S$ in a uniform way, where S is the set of non-eliminated experts. Then the tests $\Delta'_{ji} \leq 0$ and $\Delta'_{ij} \leq 0$ are performed, where Δ'_{ij} is defined by (5). If i or j fail the test, which means that it is a suboptimal expert, it is eliminated from S.

Finally, when the algorithm is halted, depending on the number of allocated samples, we choose either an empirical risk minimizer over the non-eliminated experts or the mean of two experts from S that are distant enough. This rule allows our algorithm's output to enjoy the best of converge rates of the two methods.

We introduce the following notations: In round t, denote $T_{ij}(t)$ the number of samples where predictions of experts i and j were jointly queried and $T_i(t)$ the number of rounds where the prediction of expert i was queried. Denote $\hat{R}_{ij}(j, t)$ the empirical loss of expert i calculated using only the $T_{ij}(t)$ samples queried for (i, j) jointly. We define $\alpha_{ij}(t, \delta) := \sqrt{\frac{\log(4K\delta^{-1})}{T_{ij}(t)}}$ if $T_{ij}(t) > 0$ and $\alpha_{ij}(t) = \infty$ otherwise. Let $\hat{d}_{ij}(t)$ be the empirical L_2 distance between experts i and j based on the $T_{ij}(t)$ queried samples. Denote $\delta_t := \delta/(t(t+1))$. For $i, j \in [\![K]\!]$ we define:

$$\Delta'_{ij}(t, \delta) := \hat{R}_{ij}(j, t) - \hat{R}_{ij}(i, t) - 6 \max\Big\{ L\alpha_{ij}(t, \delta_t)\hat{d}_{ij}(t), B\alpha^2_{ij}(t, \delta_t) \Big\}. \tag{5}$$

Algorithm 2 Two-point feedback

Input δ, L and ρ.
Initialization: $S \leftarrow [\![K]\!]$.
for $T = 1, 2, \ldots$ **do**
  Let $(i, j) \in \operatorname{Arg\,Min}_{(u,v)\in S\times S} T_{uv}$.
  Query the advice of experts i and j and update the corresponding quantities.
  For all u, v: If $\Delta'_{uv} > 0$: $S \leftarrow S \setminus \{v\}$.
end for
On interrupt: Let $\hat{k} \in S$ and let $\hat{l} \leftarrow \operatorname*{argmax}_{j\in S} \hat{d}_{\hat{k}j}$.
Let $\hat{q}$ denote the empirical risk minimizer on S.
if $T_{\hat{k}\hat{l}} > \sqrt{\log(KT\delta^{-1})T_{\hat{q}}}$ **then**
  Return $\frac{1}{2}\big(F_{\hat{k}} + F_{\hat{l}}\big)$.
else
  Return $F_{\hat{q}}$.
end if

Our first result in this setting is an empirical bound. At any interruption time, it gives a bound on the excess risk, only depending on quantities available to the user, using the number of queries resulting from the querying strategy in Algorithm 2. We then use a worst-case bound on these quantities to develop an instance independent bound in Corollary 5.2.

Theorem 5.1. *(Empirical bound) Suppose Assumption 1 holds. Let $T \geq 2K^2$, and denote $\hat{g}$ the output of Algorithm 2 with inputs (δ, L, ρ) in round T. Then with probability at least $1 - 3\delta$:*

$$R(\hat{g}) \leq R^* + c\, B \min\left\{ \frac{\log\big(TK\delta^{-1}\big)}{T_{\hat{k}\hat{l}}(T)}, \sqrt{\frac{\log(TK\delta^{-1})}{T_{\hat{q}}(T)}} \right\}, \tag{6}$$

where $\hat{k}, \hat{l}$ and $\hat{q}$ are the experts in Algorithm 2 and c is an absolute constant.

Proof Sketch of Theorem 5.1 We start by noting that when running Algorithm 2, the optimal experts $\mathcal{S}^* = \operatorname{Arg\,Min}_{i \in [\![K]\!]} R(F_i)$ are never eliminated with high probability (Lemma S-5). This shows in particular, that when the procedure is terminated, we have $\mathcal{S}^* \subseteq S_T$, where S_T is the set of non-eliminated experts at round T.

Then we show the following key result: in each round $t \leq T$, for any expert $i \in S_t$, let $j \in \operatorname{Arg\,Max}_{l \in S_t} \hat{d}_{il}(t)$, we have with probability at least $1 - \delta$:

$$R\left(\frac{F_i + F_j}{2}\right) \leq R^* + cB\, \frac{\log(K\delta_t^{-1})}{T_{ij}(t)}.$$

For the second bound, recall that i^* belongs to S_T with high probability. Therefore, performing an empirical risk minimization over the set of non-eliminated experts leads to the bound $\sqrt{\frac{\log(KT\delta^{-1})}{T_q(T)}}$, through a simple concentration argument using Hoeffding's inequality.

Corollary 5.2. *(Instance independent bound) Suppose assumption 1 holds. Let $T \geq 2K^2$, and denote $\hat{g}$ the output of Algorithm 2 with inputs (δ, L, ρ) in round T. Then with probability at least $1 - 3\delta$:*

$$R(\hat{g}) \leq R^* + c\, B \min\left\{ \frac{K^2 \log\big(TK\delta^{-1}\big)}{T}, \sqrt{\frac{K \log(TK\delta^{-1})}{T}} \right\},$$

where c is an absolute constant.

Proof. We develop an elementary bound on $T_{\hat{k}\hat{l}}$ and $T_{\hat{q}}$, then we inject these bounds into inequality (6).

Note that: $\hat{q}, i^* \in S_T$, hence $T_{\hat{q}}(T), T_{i^*}(T) \geq \frac{T}{2K}$. Moreover, we have:

$$T_{\hat{k}\hat{l}}(T) \geq \frac{T}{K^2}.$$

Using inequality (6), we obtain the result. □

Remarks. *Observe that in all the considered settings (full information, budgeted and limited advice), the number of jointly sampled pairs (F_i, F_j) to attain an excess risk of $\mathcal{O}(\epsilon)$ is of the order of $\mathcal{O}(K^2/\epsilon)$. Being able to ask a set of m experts simultaneously in a training round allows to sample $m(m-1)/2$ pairs for a query cost of m: this is the advantage of the budgeted setting, while we have to query each pair in succession under the strict $m = 2$ constraint, resulting in a higher cost overall.*

Theorem 5.3. *(Instance dependent bound) Suppose Assumption 1 holds. Let $\hat{g}$ denote the output of Algorithm 2 with input (δ, L, ρ) and T denote the total number of rounds. Let $\epsilon > 0$, if :*

$$T \geq 578\, C_\epsilon \log\big(\delta^{-1} C_\epsilon\big),$$

where

$$C_\epsilon := K \sum_{i \in \mathcal{S}_\epsilon^c} \Lambda_i + 2|\mathcal{S}_\epsilon|^2 \min\left\{ \frac{1}{\epsilon}, \Lambda^* \right\},$$

where $\Lambda^* := \max_{i:\Lambda_i < +\infty} \Lambda_i$, then, with probability at least $1 - \delta$:

$$R(\hat{g}) \leq R^* + cB\, \epsilon,$$

where c is an absolute constant.

Remarks. *If the algorithm is allowed to query $m > 2$ expert advices per round, then it can be modified to attain an improved excess risk. We present this extension in Section S-8 in the supplemental, and prove that it leads to a rate of $\mathcal{O}\Big(\frac{(K/m)^2}{T}\log(KT/\delta)\Big)$, which interpolates for intermediate values of m.*

Proof Sketch of Theorem 5.3 First, we develop instance-dependent upper and lower bound for $T_{ij}(t)$, for any $i, j \in [\![K]\!]$ such that: $R(F_i) \neq R(F_j)$. To do this we introduce the following lemma (see Lemma S-7 in the supplemental):

Lemma 5.4. *Let $i, j \in [\![K]\!]$ such that $R(F_i) \neq R(F_j)$. With probability at least $1 - 4\delta$, for all $t \geq 1$, if*

$$T_{ij}(t) \geq 289 \log\big(K\delta_t^{-1}\big) \max\left\{ \frac{L^2 d_{ij}^2}{|R(F_i) - R(F_j)|^2}; \frac{B}{|R(F_i) - R(F_j)|} \right\},$$

then we have either $\Delta'_{ij} > 0$ or $\Delta'_{ji} > 0$; furthermore, if

$$T_{ij}(t) \leq 3 \log\big(K\delta_t^{-1}\big) \max\left\{ \frac{L^2 d_{ij}^2}{|R(F_i) - R(F_j)|^2}; \frac{B}{|R(F_i) - R(F_j)|} \right\},$$

then we have: $\Delta'_{ij} \leq 0$ and $\Delta'_{ji} \leq 0$.

This lemma gives in particular an upper bound on the number of allocations needed for an expert i to be eliminated by an optimal expert i^* (i.e. to fail the test $\Delta_{ii^*} \leq 0$). Then, we derive a bound on the number of rounds T_ϵ required to eliminate all the experts in $\mathcal{S}_\epsilon^c$ and we conclude by showing that $T - T_\epsilon$ is large enough to ensure that the experts $\hat{k}$ and $\hat{l}$ in algorithm 2 satisfy $T_{\hat{k}\hat{l}} > 1/\epsilon$ with high probability.

6 Lower Bounds for $m = 1$ or $p = 1$

This section considers the case where the agent is restricted to selecting one expert at the end of the procedure ($p = 1$), and the case where the learner is restricted to see only one feedback per round ($m = 1$). We show that in either case it is impossible to do better than an excess risk $\mathcal{O}\big(1/\sqrt{T}\big)$ in deviation.

Lemma 6.1 is a direct consequence of a more general lower bound in [18], which proved that if the closure of the experts class is non-convex, and a single expert must be picked at the end ("proper" learning rule), then even under full information access during training the best achievable rate with high probability is $\mathcal{O}\big(1/\sqrt{T}\big)$.

Lemma 6.1. *($p = 1$) Consider the squared loss function. For $K = m = 2$ and $p = 1$, for any $T > 0$, and for any convex combination of the experts $\hat{g}$ output after T training rounds, there exists a probability distribution for experts $\{F_1, F_2\}$ and target variable Y (all bounded by 1) such that, with probability at least 0.1,*

$$\hat{R}_T(\hat{g}) - R^* \geq \frac{c_1}{\sqrt{T}},$$

where $c_1 > 0$ is an absolute constant.

The second result shows that the same lower bound holds for the bandit feedback ($m = 1$) setting, even if the learner is allowed to predict using a convex combination of all the experts at the end. To the best of our knowledge, this is the first lower bound for deviations in this setting.

Lemma 6.2. *($m = 1$) Consider the squared loss function. For $K = p = 2$, and $m = 1$, for any $T > 0$, for any convex combination of the experts $\hat{g}$ output after T training rounds, there exists a probability distribution for experts $\{F_1, F_2\}$ and target variable Y (all bounded by 1) such that with probability at least 0.1,*

$$\hat{R}_T(\hat{g}) - R^* \geq \frac{1}{2\sqrt{T}}.$$

7 Conclusion

We discussed the impact of restricted access to information in generalization error minimization with respect to the best expert. As many classical methods, such as progressive mixture rules (and randomized versions thereof) are deviation suboptimal, we proposed a new procedure achieving fast rates with high probability. We focused on the global budget setting, where a constraint on the total number of expert queries is made, and the local budget, where a limited number of expert advices are shown per round. Moreover, we proved fast rates are impossible to achieve if the agent is allowed to see just one expert advice per round or choose just one expert for prediction.

An interesting future direction is allowing experts to learn from data during the process. In this case, the i.i.d. assumption on the loss sequence is dropped, which necessitates deriving a new concentration for the key quantities.

Acknowledgements

We acknowledge support from the Agence Nationale de la Recherche (ANR), ANR-19-CHIA-0021-01 “BiSCottE”; and the Franco-German University (UFA) through the binational Doktorandenkolleg CDFA 01-18.

References

[1] A. Agarwal, O. Dekel, and L. Xiao. Optimal algorithms for online convex optimization with multi-point bandit feedback. In *COLT*, pages 28–40, 2010.

[2] J.-Y. Audibert. Progressive mixture rules are deviation suboptimal. In J. Platt, D. Koller, Y. Singer, and S. Roweis, editors, *Advances in Neural Information Processing Systems*, volume 20, 2008.

[3] J.-Y. Audibert. Progressive mixture rules are deviation suboptimal / Supplemental "Proof of the optimality of the empirical star algorithm". In J. Platt, D. Koller, Y. Singer, and S. Roweis, editors, *Advances in Neural Information Processing Systems*, volume 20, 2008.

[4] J.-Y. Audibert. Fast learning rates in statistical inference through aggregation. *The Annals of Statistics*, 37(4):1591–1646, 2009.

[5] J.-Y. Audibert and S. Bubeck. Regret bounds and minimax policies under partial monitoring. *The Journal of Machine Learning Research*, 11:2785–2836, 2010.

[6] S. Ben-David and E. Dichterman. Learning with restricted focus of attention. *Journal of Computer and System Sciences*, 56(3):277–298, 1998.

[7] O. Catoni. A mixture approach to universal model selection. Technical Report LMENS-97-30, Ecole Normale Supérieure, 1997.

[8] N. Cesa-Bianchi, A. Conconi, and C. Gentile. On the generalization ability of on-line learning algorithms. *IEEE Transactions on Information Theory*, 50(9):2050–2057, 2004.

[9] N. Cesa-Bianchi and G. Lugosi. *Prediction, learning, and games*. Cambridge university press, 2006.

[10] N. Cesa-Bianchi, S. Shalev-Shwartz, and O. Shamir. Efficient learning with partially observed attributes. *Journal of Machine Learning Research*, 12(10), 2011.

[11] S. Gerchinovitz and T. Lattimore. Refined lower bounds for adversarial bandits. In *Proceedings of the 30th International Conference on Neural Information Processing Systems*, pages 1198–1206, 2016.

[12] S. Guha and K. Munagala. Approximation algorithms for budgeted learning problems. In *Proceedings of the thirty-ninth annual ACM symposium on Theory of computing*, pages 104–113, 2007.

[13] D. Haussler, J. Kivinen, and M. K. Warmuth. Sequential prediction of individual sequences under general loss functions. *IEEE Transactions on Information Theory*, 44(5):1906–1925, 1998.

[14] E. Hazan and T. Koren. Optimal algorithms for ridge and lasso regression with partially observed attributes. *arXiv preprint arXiv:1108.4559*, 2011.

[15] E. Kaufmann, O. Cappé, and A. Garivier. On the complexity of best-arm identification in multi-armed bandit models. *The Journal of Machine Learning Research*, 17(1):1–42, 2016.

[16] G. Lecué. Suboptimality of penalized empirical risk minimization in classification. In *International Conference on Computational Learning Theory*, pages 142–156. Springer, 2007.

[17] G. Lecué and S. Mendelson. Aggregation via empirical risk minimization. *Probability theory and related fields*, 145(3-4):591–613, 2009.

[18] W. S. Lee, P. Bartlett, and R. Williamson. The importance of convexity in learning with squared loss. *IEEE Transactions on Information Theory*, 44(5):1974–1980, 1998.

[19] O. Madani, D. J. Lizotte, and R. Greiner. Active model selection. In *Proceedings of the 20th conference on Uncertainty in artificial intelligence*, pages 357–365, 2004.

[20] S. Mannor and O. Shamir. From bandits to experts: on the value of side-observations. In *Proceedings of the 24th International Conference on Neural Information Processing Systems*, pages 684–692, 2011.

[21] Y. Seldin, P. Bartlett, K. Crammer, and Y. Abbasi-Yadkori. Prediction with limited advice and multiarmed bandits with paid observations. In *International Conference on Machine Learning*, pages 280–287. PMLR, 2014.

[22] O. Shamir. An optimal algorithm for bandit and zero-order convex optimization with two-point feedback. *The Journal of Machine Learning Research*, 18(1):1703–1713, 2017.

[23] T. S. Thune and Y. Seldin. Adaptation to easy data in prediction with limited advice. In *Proceedings of the 32nd International Conference on Neural Information Processing Systems*, pages 2914–2923, 2018.

[24] Y. Xia, T. Qin, W. Ma, N. Yu, and T.-Y. Liu. Budgeted multi-armed bandits with multiple plays. In *IJCAI*, pages 2210–2216, 2016.

[25] Y. Yang and A. Barron. Information-theoretic determination of minimax rates ofconvergence. *The Annals of Statistics*, 27(5):1564 – 1599, 1999.

[26] D. Zhou and C. Tomlin. Budget-constrained multi-armed bandits with multiple plays. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 32, 2018.

Supplementary Material for: Fast rates for prediction with limited expert advice

El Mehdi Saad[1], **Gilles Blanchard**[1,2]
[1]Laboratoire de Mathématiques d'Orsay, CNRS, Université Paris-Saclay; [2]Inria

1 Notation

The following notation pertains to all the considered algorithms, where t is a given training round:

- Let $\mathcal{T}_i(t)$ denote the set of training round indices where the advice of expert i was queried and let $T_i(t) := |\mathcal{T}_i(t)|$.
- Let $\mathcal{T}_{ij}(t)$ denote the set of training round indices where the advice of experts i and j where jointly queried and let $T_{ij}(t) := |\mathcal{T}_{ij}(t)|$.
- Let $\hat{R}_{ij}(j,t)$ denote the empirical loss of expert j calculated using only the $T_{ij}(t)$ samples queried for (i,j) jointly:

$$\hat{R}_{ij}(j,t) := \frac{1}{T_{ij}(t)} \sum_{s \in \mathcal{T}_{ij}(t)} l(F_{j,s}, Y_s).$$

- $\hat{R}_i(t)$ denote the empirical loss of expert i calculated using the $T_i(t)$ queried samples:

$$\hat{R}_i(t) := \frac{1}{T_i(t)} \sum_{s \in \mathcal{T}_i(t)} l(F_{i,s}, Y_s).$$

- Define $\alpha_{ij}(t,\delta) := \sqrt{\frac{\log(4K\delta^{-1})}{T_{ij}(t)}}$ if $T_{ij}(t) > 0$ and $\alpha_{ij}(t) = \infty$ otherwise.
- Define $\alpha_i(t,\delta) := \sqrt{\frac{\log(4K\delta^{-1})}{T_i(t)}}$ if $T_i(t) > 0$ and $\alpha_i(t) = \infty$ otherwise.
- Let $\hat{d}_{ij}(t)$ denote the empirical L_2 distance between experts i and j based on the $T_{ij}(t)$ queried samples:

$$\hat{d}^2_{ij}(t) := \frac{1}{T_{ij}(t)} \sum_{s \in \mathcal{T}_{ij}(t)} (F_{i,s} - F_{j,s})^2.$$

- Define $\Delta'_{ij}(t,\delta) := \hat{R}_{ij}(j,t) - \hat{R}_{ij}(i,t) - 6\alpha_{ij}(t,\delta) \max\Big\{L\hat{d}_{ij}(t), B\alpha_{ij}(t,\delta)\Big\}$.
- Let d_{ij} denote the L_2 distance between experts i and j:

$$d_{ij} := \mathbb{E}\Big[(F_i - F_j)^2\Big].$$

- We denote $R(.)$ the expected risk function: $R(.) = \mathbb{E}[l(.,Y)]$, and define $R_i = R(F_i)$ for $i \in [\![K]\!]$.

35th Conference on Neural Information Processing Systems (NeurIPS 2021).

2 Some preliminary results

The lemma below shows that for a set $\mathcal{Y}$ and a convex set $\mathcal{X} \subseteq \mathbb{R}^d$, if there exists a function $l : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$ that is Lipschitz and strongly convex on its first argument, then the function l and the set $\mathcal{X}$ are bounded.

Lemma 1. Let $\mathcal{X} \subseteq \mathbb{R}^d$ be a non-empty convex set, let $\mathcal{Y}$ be an arbitrary set and $l : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$ be a function such that for all $y \in \mathcal{Y} : l(., y)$ is L-Lipschitz and ρ-strongly convex, then we have:

- $\sup_{x,x' \in \mathcal{X}} \|x - x'\| \leq \frac{B}{L} = 8\frac{L}{\rho^2}$.
- $\sup_{x,x' \in \mathcal{X}, y \in \mathcal{Y}} |l(x, y) - l(x', y)| \leq B := 8\frac{L^2}{\rho^2}$

Proof. Let $y \in \mathcal{Y}$ and $x_0, x \in \mathcal{X}$, using the ρ-strong convexity of $l(., y)$ we have:

$$l\left(\frac{x + x_0}{2}, y\right) - \frac{\rho^2}{2}\left\|\frac{x + x_0}{2}\right\|^2 \leq \frac{1}{2}\left(l(x_0, y) - \frac{\rho^2}{2}\|x_0\|^2\right) + \frac{1}{2}\left(l(x, y) - \frac{\rho^2}{2}\|x\|^2\right)$$

Which implies:

$$\frac{\rho^2}{2}\left(\frac{1}{4}\|x_0 + x\|^2 - \frac{1}{2}\|x_0\|^2 - \frac{1}{2}\|x\|^2\right) \leq l\left(\frac{x + x_0}{2}, y\right) - \frac{l(x, y) + l(x_0, y)}{2}.$$

Using the parallelogram law and the assumption that l is L-Lipschitz we have:

$$\frac{\rho^2}{8}\|x - x_0\|^2 \leq L\|x - x_0\|,$$

which proves that $\text{diam}(\mathcal{X}) \leq 8\frac{L}{\rho^2}$. Now using the assumption that $l(., y)$ is L-Lipschitz, we have:

$$\begin{aligned} |l(x, y) - l(x_0, y)| &\leq L\|x - x_0\| \\ &\leq 8\frac{L^2}{\rho^2}, \end{aligned}$$

which proves the second claim. □

For any $y \in \mathcal{Y}$, let $l^*(y) = \min_{x \in \mathcal{X}} l(x, y)$, which exists since l is continuous in x and $\mathcal{X}$ is a closed bounded set by the previous lemma, and let $\widetilde{l}(x, y) := l(x, y) - l^*(y)$. By the previous lemma, $\widetilde{l}(x, y) \in [0, B]$; also, note that the proposed algorithms remain unchanged if we replace the loss l by $\widetilde{l}$, since the algorithms only depend on loss differences for different predictions x, x' and the same y. Similarly, the excess loss of any predictor remains unchanged when replacing l by $\widetilde{l}$. Therefore, without loss of generality we can assume that the loss function always takes values in $[0, B]$, which we do for the remainder of the paper.

The following lemma is technical, it will be used in the proof of the instance dependent bound (Theorem M-5.3).

Lemma 2. Let $x \geq 1, c \in (0, 1)$ and $y > 0$ such that:

$$\frac{\log(x/c)}{x} > y. \tag{1}$$

Then:

$$x < \frac{2\log\left(\frac{1}{cy}\right)}{y}.$$

Proof. Inequality (1) implies

$$x < \frac{\log(x/c)}{y},$$

and further

$$\log(x/c) < \log(1/yc) + \log\log(x/c) \leq \log(1/yc) + \frac{1}{2}\log(x/c),$$

since it can be easily checked that $\log(t) \leq t/2$ for all $t > 0$. Solving and plugging back into the previous display leads to the claim. □

3 Some concentration results

In this section, we present concentration inequalities for the key quantities used in our analysis. Recall that Lemma 1 shows that under assumption M-1, without loss of generality we can assume that the loss function takes values in $[0, B]$, $B := 8L^2/\rho^2$.

The following lemma gives the main concentration inequalities we need:

Lemma 3. Suppose Assumption M-1 holds. For any integer $t \geq 1$, and $\delta \in [0, 1]$, with probability at least $1 - 3\delta$, for all $i, j \in [\![K]\!]$:

$$\left|\left(\hat{R}_{ij}(i,t) - \hat{R}_{ij}(j,t)\right) - (R_i - R_j)\right| \leq \sqrt{2}L\ \hat{d}_{ij}\ \alpha_{ij}(t,\delta) + 3B\ \alpha_{ij}^2(t,\delta)$$

$$\left|\hat{d}_{ij}^2 - d_{ij}^2\right| \leq \max\left\{2\frac{B}{L}\ \alpha_{ij}(t,\delta)\ d_{ij}\ ;\ 6\left(\frac{B}{L}\right)^2\ \alpha_{ij}^2(t,\delta)\right\}$$

$$\left|\hat{R}_i(t) - R_i\right| \leq 2B\alpha_i(t,\delta).$$

Proof. The first inequality is a direct consequence of the empirical Bernstein inequality (Theorem 4 in [3]). Recall that l is L-Lipschitz in its first argument. Hence, we have the following bound on the empirical variance of the variable: $l(F_i, Y) - l(F_j, Y)$.

$$\begin{aligned}\widehat{\mathrm{Var}}[l(F_i,Y) - l(F_j,Y)] &:= \frac{2}{T_{ij}(t)(T_{ij}(t)-1)} \sum_{u,v \in \mathcal{T}_{ij}(t)} \left(l(F_{i,u},Y_u) - l(F_{j,u},Y_u) - l(F_{i,v},Y_v) + l(F_{j,v},Y_v)\right)^2 \\ &\leq \frac{1}{T_{ij}(t)} \sum_{u \in \mathcal{T}_{ij}(t)} \left(l(F_{i,u},Y_u) - l(F_{j,u},Y_u)\right)^2 \\ &\leq L^2\ \hat{d}_{ij}^2.\end{aligned}$$

The second inequality is a consequence of Bernstein inequality applied to $\hat{d}_{ij}^2$, we used the following bound on the variance of the variable $(F_i - F_j)^2$:

$$\begin{aligned}\mathrm{Var}\left[(F_i - F_j)^2\right] &\leq \mathbb{E}\left[\|F_i - F_j\|^4\right] \\ &\leq \sup_{i,j \in [K]} \|F_i - F_j\|^2 \mathbb{E}\left[\|F_i - F_j\|^2\right] \\ &\leq \left(\frac{B}{L}\right)^2 d_{ij}^2.\end{aligned}$$

Finally, the last inequality stems from Hoeffding's inequality. □

Corollary 4. Let $T > 0$ be fixed. In the full information case ($m = K$), with probability at least $1 - 2\delta$, it holds:

$$\text{For all } i, j \in [\![K]\!]: \qquad \Delta_{ij} \leq (R_j - R_i) \leq \Delta_{ij} + 32\alpha \max(Ld_{ij}, B\alpha). \tag{2}$$

Proof. In the full information case, since all experts are queried at each round we have $T_{ij}(T) = T_i(T) = T$ and $\alpha_{ij}(T,\delta) = \alpha(T,\delta) = \alpha$ for all i, j. Applying Lemma 3 in that setting, using the first inequality we obtain that with probability at least $1 - 3\delta$:

$$\Delta_{ij} \leq \left(\hat{R}(i,T) - \hat{R}(j,T)\right) - \sqrt{2}L\hat{d}_{ij}\alpha - 3B\alpha^2 \leq R_i - R_j,$$

giving the first inequality in (2); and

$$R_i - R_j \leq \left(\hat{R}(i,T) - \hat{R}(j,T)\right) + \sqrt{2}L\hat{d}_{ij}\alpha + 3B\alpha^2 \leq \Delta_{ij} + 9\alpha L\hat{d}_{ij} + 9B\alpha^2. \tag{3}$$

From the second inequality in Lemma 3 we get, putting $\beta := B/L$:

$$\begin{aligned}\hat{d}_{ij}^2 - d_{ij}^2 &\leq \max\left\{2\beta\alpha d_{ij}, 6\beta^2\alpha^2\right\} \\ &\leq \max\left\{6\beta^2\alpha^2 + \frac{1}{6}d_{ij}^2, 6\beta^2\alpha^2\right\} \\ &\leq 6\beta^2\alpha^2 + \frac{1}{6}d_{ij}^2,\end{aligned}$$

from which we deduce $\hat{d}_{ij}^2 \leq 12\alpha \max(\beta^2\alpha^2, d_{ij}^2)$. Taking square roots and plugging into (3), we obtain the claim. □

For $t \geq 1$, define: $\delta_t := \frac{\delta}{t(t+1)}$. Define the event $\mathcal{A}$:

$$(\mathcal{A}) : \forall t \geq 1, \forall\, i,j \in [\![K]\!] : \begin{cases} \left|\left(\hat{R}_{ij}(i,t) - \hat{R}_{ij}(j,t)\right) - (R_i - R_j)\right| \leq 3\max\left\{L\hat{d}_{ij}\ \alpha_{ij}(t,\delta_t); B\alpha_{ij}^2(t,\delta_t)\right\} & \text{(4a)} \\ \left|\hat{R}_i(t) - R_i\right| \leq 2B\ \alpha_i(t,\delta_t) & \text{(4b)} \\ \hat{d}_{ij}^2 \leq 12\max\left\{d_{ij}^2; \left(\frac{B}{L}\right)^2 \alpha_{ij}^2(t,\delta_t)\right\} & \text{(4c)} \\ d_{ij}^2 \leq 12\max\left\{\hat{d}_{ij}^2; \left(\frac{B}{L}\right)^2 \alpha_{ij}^2(t,\delta_t)\right\} & \text{(4d)} \end{cases}$$

Using a union bound over $t \geq 1$ and $i,j \in [\![K]\!]$, we have: $\mathbb{P}(\mathcal{A}) \geq 1 - 4\delta$.

4 Proof of Theorem M-5.1 and Corollary M-5.2

Let $t \geq 1$, denote by S_t the set of non-eliminated experts in Algorithm M-2 at round t. The lemma below shows that conditionally to event $\mathcal{A}$, the best experts $\mathcal{S}^*$ are never eliminated.

Lemma 5. If $\mathcal{A}$ defined in (4) holds, $\forall t \geq 1$ we have: $\mathcal{S}^* \subseteq S_t$, where we recall $\mathcal{S}^* := \operatorname{Arg\,Min}_{i \in [\![K]\!]} R(F_i)$.

Proof. Let $t \geq 1$, assume for the sake of contradiction that: $i^* \in \mathcal{S}^*$ but $i^* \notin S_t$. Then, at some point, i^* was eliminated by an expert j. More specifically: $\exists s \in [\![t]\!]$, $\exists j \in [\![K]\!] \setminus \{i^*\}$, such that $\Delta'_{ji^*}(t,\delta_t) > 0$. It follows by definition of Δ'_{ji^*} that:

$$\hat{R}_{ji^*}(i^*,s) > \hat{R}_{ji^*}(j,s) + 6\max\left\{L\alpha_{ji^*}(s,\delta_s)\hat{d}_{ji^*}, B\alpha_{ji^*}^2(s,\delta_s)\right\}$$

which contradicts (4a) since we have: $R^* \leq R_j$. □

The lemma below gives a high probability deviation rate on the excess of any expert in S_t when combined with an appropriate expert. Recall that for $i \in [\![K]\!]$: $R_i = R(F_i)$.

Lemma 6. If event $\mathcal{A}$ defined in (4) holds, $\forall t \geq 1$, for all $i \in S_t$, let $j \in \operatorname{argmax}_{l \in S_t} \hat{d}_{il}(t)$, then we have:

$$R\left(\frac{F_i + F_j}{2}\right) \leq R^* + c\,B\frac{\log(K\delta_t^{-1})}{T_{ij}(t)},$$

where c is an absolute constant.

Proof. Suppose that $\mathcal{A}$ is true. Let $t \geq 1$, $i \in S_t$ and $i^* \in \mathcal{S}^*$. Let $j \in \operatorname{argmax}_{S_t} \hat{d}_{il}$.

Lemma 5 shows that : $i^* \in S_t$, we therefore have by construction of Algorithm M-2:

$$\hat{R}_{ij}(j,t) \leq \hat{R}_{ij}(i,t) + 6\max\left\{L\alpha_{ij}(t,\delta_t)\hat{d}_{ij}(t), B\alpha_{ij}^2(t,\delta_t)\right\}$$
$$\hat{R}_{ii^*}(i,t) \leq \hat{R}_{ii^*}(i^*,t) + 6\max\left\{L\alpha_{ii^*}(t,\delta_t)\hat{d}_{ii^*}(t), B\alpha_{ii^*}^2(t,\delta_t)\right\}.$$

Using inequalities (4a) for (i,j) and (i,i^*) respectively and $\hat{d}_{ii^*}(t) \leq \hat{d}_{ij}(t)$, we have:

$$R_j \leq R_i + 9\max\left\{L\alpha_{ij}(t,\delta_t)\hat{d}_{ij}(t), B\alpha_{ij}^2(t,\delta_t)\right\} \quad (5)$$
$$R_i \leq R_{i^*} + 9\max\left\{L\alpha_{ii^*}(t,\delta_t)\hat{d}_{ij}(t), B\alpha_{ii^*}^2(t,\delta_t)\right\}. \quad (6)$$

We have:

$$\begin{aligned}
R\Big(\frac{F_i+F_j}{2}\Big) &\le \frac{1}{2}\Big(R_i - \frac{\rho^2}{2}\mathbb{E}\big[F_i^2\big]\Big) + \frac{1}{2}\Big(R_j - \frac{\rho^2}{2}\mathbb{E}\big[F_j^2\big]\Big) + \frac{\rho^2}{2}\mathbb{E}\left[\left(\frac{F_i+F_j}{2}\right)^2\right] \\
&= \frac{1}{2}R_i + \frac{1}{2}R_j - \frac{\rho^2}{8}\Big(2\mathbb{E}\big[F_i^2\big] + 2\mathbb{E}\big[F_j^2\big] - \mathbb{E}\big[(F_i+F_j)^2\big]\Big) \\
&= \frac{1}{2}R_i + \frac{1}{2}R_j - \frac{\rho^2}{8}d_{ij}^2 \\
&\le \frac{1}{2}R_i + \frac{1}{2}R_i + \frac{9}{2}\max\Big\{L\alpha_{ij}(t,\delta_t)\hat{d}_{ij}(t), B\alpha_{ij}^2(t,\delta_t)\Big\} - \frac{\rho^2}{8}d_{ij}^2 \\
&= R_i + \frac{9}{2}\max\Big\{L\alpha_{ij}(t,\delta_t)\hat{d}_{ij}(t), B\alpha_{ij}^2(t,\delta_t)\Big\} - \frac{\rho^2}{8}d_{ij}^2 \\
&\le R^* + \frac{27}{2}\max\Big\{L\alpha_{ij}(t,\delta_t)\hat{d}_{ij}(t), B\alpha_{ij}^2(t,\delta_t)\Big\} - \frac{\rho^2}{8}d_{ij}^2.
\end{aligned}$$

We used the strong convexity of R in the first inequality and we injected (5) to bound $R(F_j)$ in the fourth line and (6) to bound $R(F_i)$ in the last line. Now we use inequality (4b) for (i,j) and obtain:

$$\begin{aligned}
R\Big(\frac{F_i+F_j}{2}\Big) - R^* &\le 162\max\big\{L\alpha_{ij}(t,\delta_t)d_{ij}, B\alpha_{ij}^2(t,\delta_t)\big\} - \frac{\rho^2}{8}d_{ij}^2 \\
&\le c\,B\alpha_{ij}^2(t,\delta_t) \\
&\le c\,B\alpha_{ij}^2(t,\delta_t),
\end{aligned}$$

where c is an absolute constant. In the final step, we upper bounded the right-hand-side of the first inequality with a parabolic function in d_{ij}, then we replaced d_{ij} with the expression achieving the maximum (recall that $B := 8(L/\rho)^2$).

□

Proof of Theorem M-5.1. Let $T \ge 2K^2$, when Algorithm M-2 is halted at T. Let $\hat{k} \in S_T$ and $\hat{l} \in \text{argmax}_{j\in S_T}\hat{d}_{\hat{k}j}(T)$.

Let $\hat{q}$ denote the empirical risk minimizer on S_T:

$$\hat{q} \in \operatorname*{Arg\,Min}_{j\in S_T} \hat{R}_j(T).$$

We consider two cases. If $T_{\hat{k}\hat{l}}(T) > \sqrt{T_{\hat{q}}(T)\log\big(K\delta_T^{-1}\big)}$, then the output of Algorithm M-2 is $\frac{F_{\hat{k}}+F_{\hat{l}}}{2}$ and we can apply the bound of Lemma 6.

If $T_{\hat{k}\hat{l}}(T) \le \sqrt{T_{\hat{q}}(T)\log\big(K\delta_T^{-1}\big)}$, then the output of Algorithm M-2 is $F_{\hat{q}}$. We have:

$$\begin{aligned}
R_{\hat{q}} - R_{i^*} &= R_{\hat{q}} - \hat{R}_{\hat{q}}(T) + \hat{R}_{\hat{q}}(T) - \hat{R}_{i^*}(T) + \hat{R}_{i^*}(T) - R_{i^*} \\
&\le 2B\sqrt{\frac{\log\big(K\delta_T^{-1}\big)}{T_{\hat{q}}(T)}} + 2B\sqrt{\frac{\log\big(K\delta_T^{-1}\big)}{T_{i^*}(T)}} \\
&\le 2B\sqrt{\frac{\log\big(K\delta_T^{-1}\big)}{T_{\hat{q}}(T)}} + 2B\sqrt{\frac{\log\big(K\delta_T^{-1}\big)}{T_{\hat{q}}(T)-K}} \\
&\le 5B\sqrt{\frac{\log\big(K\delta_T^{-1}\big)}{T_{\hat{q}}(T)}},
\end{aligned}$$

where we used inequalities (4c) for $\hat{q}$ and i^*, and the fact that the allocation strategy leads to $|T_{i^*}(T) - T_{\hat{q}}(T)| \le K$ and $T_i(T) > 2K$ for all i.

As a conclusion we have:

$$R(\hat{g}) - R_{i^*} \le c\,B\min\left\{\frac{\log(KT\delta^{-1})}{T_{\hat{k}\hat{l}}(T)}; \sqrt{\frac{\log(KT\delta^{-1})}{T_{\hat{q}}(T)}}\right\}, \tag{7}$$

where c is an absolute constant.

5 Proof of Theorem M-5.3

In this section, we prove instance dependent bounds on the number of rounds required to achieve a risk at least as good as the best expert up to $\epsilon > 0$.

The following lemma gives an instance dependent upper and lower bound on the quantities $T_{ij}(t)$, for $i, j \in [\![K]\!]$.

Lemma 7. Let $i, j \in [\![K]\!]$ such that $R_i \neq R_j$. If $\mathcal{A}$ holds, for all $t \geq 1$, if

$$T_{ij}(t) \geq 289 \log\big(K\delta_t^{-1}\big) \max\left\{ \frac{L^2 d_{ij}^2}{|R_i - R_j|^2}; \frac{B}{|R_i - R_j|} \right\},$$

then we have either $\Delta'_{ij} > 0$ or $\Delta'_{ji} > 0$.

Furthermore, if

$$T_{ij}(t) \leq 3 \log\big(K\delta_t^{-1}\big) \max\left\{ \frac{L^2 d_{ij}^2}{|R_i - R_j|^2}; \frac{B}{|R_i - R_j|} \right\},$$

then we have $\Delta'_{ij} \leq 0$ and $\Delta'_{ji} \leq 0$.

Proof. We start by proving the first claim of the lemma. Let $i, j \in [\![K]\!]$ and $t \geq 1$ such that:

$$T_{ij}(t) \geq 289 \log\big(K\delta_t^{-1}\big) \max\left\{ \frac{L^2 d_{ij}^2}{|R_i - R_j|^2}; \frac{B}{|R_i - R_j|} \right\}. \tag{8}$$

Inequality (8) implies:

$$\alpha_{ij}(t, \delta_t) \leq \frac{1}{17} \min\left\{ \frac{|R_i - R_j|}{L d_{ij}}; \sqrt{\frac{|R_i - R_j|}{B}} \right\}.$$

By simple calculus, we see that:

$$17 \ \max\big\{ L\alpha_{ij}(t, \delta_t) d_{ij} \ ; \ B\alpha_{ij}^2(t, \delta_t) \big\} \leq |R_i - R_j|.$$

Now we use inequality (4a) from event $\mathcal{A}$ to upper bound $|R_i - R_j|$:

$$17 \ \max\big\{ L\alpha_{ij}(t, \delta_t) d_{ij}; B\alpha_{ij}^2(t, \delta_t) \big\} \leq \Big| \hat{R}_{ij}(i, t) - \hat{R}_{ij}(j, t) \Big| + 3 \max\Big\{ L\alpha_{ij}(t, \delta_t) \hat{d}_{ij}(t); B\alpha_{ij}^2(t, \delta_t) \Big\}. \tag{9}$$

Using inequality (4b), we have:

$$\max\left\{ \hat{d}_{ij}(t); \frac{B}{L} \alpha_{ij}(t, \delta_t) \right\} \leq 2\sqrt{3} \max\left\{ d_{ij}; \frac{B}{L} \alpha_{ij}(t, \delta_t) \right\}.$$

We plug in the inequality above in (9) and obtain:

$$6 \max\Big\{ L\alpha_{ij}(t, \delta_t) \hat{d}_{ij}(t); B\alpha_{ij}^2(t, \delta_t) \Big\} < \Big| \hat{R}_{ij}(i, t) - \hat{R}_{ij}(j, t) \Big|,$$

implying that we have either $\Delta'_{ij}(t) > 0$ or $\Delta'_{ji}(t) > 0$.

For the second claim, Let $i, j \in [\![K]\!]$ and $t \in [\![T]\!]$ such that:

$$T_{ij}(t) \leq 3 \log\big(K\delta_t^{-1}\big) \max\left\{ \frac{L^2 d_{ij}^2}{|R_i - R_j|^2}; \frac{B}{|R_i - R_j|} \right\}. \tag{10}$$

If $T_{ij}(t) = 0$, then $\Delta'_{ij} = \Delta'_{ji} = -\infty$.

Otherwise, inequality (10) implies that:

$$|R_i - R_j| \leq 3 \max\big\{ L\alpha_{ij}(t, \delta_t) d_{ij} \ ; \ B\alpha_{ij}^2(t, \delta_t) \big\}.$$

Now we use inequality (4a) from event $\mathcal{A}$ to lower bound $|R_i - R_j|$. We have:

$$\left|\hat{R}_{ij}(i,t) - \hat{R}_{ij}(j,t)\right| - 3\max\Big\{L\alpha_{ij}(t,\delta_t)\hat{d}_{ij}(t)\ ;\ B\alpha_{ij}^2(t,\delta_t)\Big\} \leq 3\max\big\{L\alpha_{ij}(t,\delta_t)d_{ij}\ ;\ B\alpha_{ij}^2(t,\delta_t)\big\}.$$

We plug in inequality (4d) to upper bound d_{ij}. We conclude that:

$$\left|\hat{R}_{ij}(i,t) - \hat{R}_{ij}(j,t)\right| \leq 6\max\Big\{L\alpha_{ij}(t,\delta_t)\hat{d}_{ij}(t); B\alpha_{ij}^2(t,\delta_t)\Big\},$$

implying that we have: $\Delta'_{ij}(t) \leq 0$ and $\Delta'_{ji}(t) \leq 0$. □

Now we turn to the proof of Theorem M-5.3. Recall the following notations: for $i \in [\![K]\!]$ define:

$$\Lambda_i := \min_{i^* \in \mathcal{S}^*} \max\left\{\frac{L^2 d_{ii^*}^2}{|R_i - R_{i^*}|^2}; \frac{B}{R_i - R_{i^*}}\right\}.$$

Denote the corresponding reordered values:

$$\Lambda_{(1)} \leq \Lambda_{(2)} \leq \cdots \leq \Lambda_{(K)} = +\infty,$$

and $\Lambda^* := \min\{\Lambda_i; \Lambda_i < +\infty\}$.

Proof of Theorem M-5.3. By Lemma 6, in order to show that $R(\hat{g}) \leq R^* + cB\epsilon$, it suffices to prove that for any $i, j \in S_T$, it holds $T_{ij}(T) \geq B\log(K\delta_T^{-1})/\epsilon$.

Let $\epsilon > 0$, define the following sequences, for $N \in [\![K-1]\!]$:

$$\begin{cases} \phi_N & := 289(K-N)^2\big(\Lambda_{(N)} - \Lambda_{(N-1)}\big)\log\big(\delta^{-1}C_\epsilon\big); \\ \tau_N & := \sum_{k=1}^N \phi_k, \end{cases}$$

where we define $\Lambda_{(0)} = 0$ and

$$C_\epsilon := K\sum_{i \in \mathcal{S}_\epsilon^c} \Lambda_i + 2|\mathcal{S}_\epsilon|^2 \min\left\{\frac{1}{\epsilon}, \Lambda^*\right\}.$$

Claim 1. *If event $\mathcal{A}$ holds, for any $N \in [\![K]\!]$ after round $\lceil\tau_N\rceil$, all experts i satisfying $\Lambda_i \leq \Lambda_{(N)}$ are necessarily eliminated.*

Proof. Recall that the number of queries required to eliminate an expert $i \in [\![K]\!]$ is upper bounded by the number of data points needed to have: $\Delta_{i^*i} > 0$ for any $i^* \in \mathcal{S}^*$, which would lead to the elimination of i by i^*.

Let i^* be an arbitrary element of $\mathcal{S}^*$. We use an induction argument, for $N = 1$ the claim is a direct consequence of the definition of τ_1 and Lemma 7. Let $N < K$ and suppose that the claim is valid for all $i \leq N$. Let j denote an expert such that $\Lambda_j = \Lambda_{(N+1)}$ and j was not eliminated before $\lceil\tau_N\rceil$. For $i \leq N$, the induction hypothesis suggests that between round $\lceil\tau_i\rceil$ and $\lceil\tau_{i+1}\rceil$ there was at most $K - i$ non-eliminated experts. Since the allocation strategy is uniform over the pairs of experts in $S \times S$, we have:

$$T_{ji^*}(\tau_{N+1}) \geq 2\sum_{i=0}^{N} \frac{\tau_{i+1} - \tau_i}{(K-i)(K-i+1)}, \tag{11}$$

where $\tau_0 = 0$. Recall that the definition of τ_i implies that:

$$\tau_{i+1} - \tau_i = 289(K-i-1)^2\log\big(C_\epsilon\delta^{-1}\big)\big(\Lambda_{(i+1)} - \Lambda_{(i)}\big). \tag{12}$$

We plug in the lower bound given in (12) into (11) to obtain:

$$T_{ji^*}(\tau_{N+1}) \geq 289\log\big(C_\epsilon\delta^{-1}\big)\Lambda_{(N+1)}.$$

Using Lemma 7 we conclude that expert j is eliminated before round τ_{N+1}, which completes the induction argument.

□

Claim 2. *We have for any $N \in [\![K]\!]$:*

$$\tau_N = 289 \log\big(C_\epsilon \delta^{-1}\big) \left(\sum_{i=1}^{N-1} (2(K-i)+1)\Lambda_{(i)} + (K-N)^2 \Lambda_{(N)} \right).$$

Proof. We have by definition of τ_N:

$$\begin{aligned}
\tau_N &= \sum_{i=1}^{N} \phi_i \\
&= \sum_{i=1}^{N} 289(K-i)^2 \big(\Lambda_{(i)} - \Lambda_{(i-1)}\big) \log\big(\delta^{-1} C_\epsilon\big) \\
&= \sum_{i=1}^{N} 289(K-i)^2 \Lambda_{(i)} \log\big(\delta^{-1} C_\epsilon\big) - \sum_{i=1}^{N} 289(K-i)^2 \Lambda_{(i-1)} \log\big(\delta^{-1} C_\epsilon\big) \\
&= 289 \log\big(\delta^{-1} C_\epsilon\big) \left(\sum_{i=1}^{N-1} (2(K-i)+1)\Lambda_{(i)} + (K-N)^2 \Lambda_{(N)} \right).
\end{aligned}$$

□

Conclusion: Let N_ϵ denote the integer satisfying (we do not consider the trivial case where all the expert have the same risk):

$$\Lambda_{(N_\epsilon)} < \frac{1}{\epsilon} < \Lambda_{(N_\epsilon+1)}.$$

Recall that we suppose that T satisfies:

$$T \geq 578 C_\epsilon \log(C_\epsilon \delta^{-1}).$$

Observe that (using Claim 2):

$$\begin{aligned}
T &\geq \tau_{N_\epsilon} + 289 \log(C_\epsilon \delta^{-1}) \left(2|\mathcal{S}_\epsilon|^2 \min\left\{ \frac{1}{\epsilon}; \Lambda^* \right\} - (K-N_\epsilon)^2 \Lambda_{(N_\epsilon)} \right) && (13) \\
&\geq \tau_{N_\epsilon} + 289 \log(C_\epsilon \delta^{-1}) \left(2|\mathcal{S}_\epsilon|^2 \min\left\{ \frac{1}{\epsilon}; \Lambda^* \right\} - |\mathcal{S}_\epsilon|^2 \Lambda^* \right) && (14) \\
&\geq \tau_{N_\epsilon} + 289 \log(C_\epsilon \delta^{-1}) |\mathcal{S}_\epsilon|^2 \min\left\{ \frac{1}{\epsilon}; \Lambda^* \right\}. && (15)
\end{aligned}$$

Claims 1 and 2 show that after $\lceil \tau_{N_\epsilon} \rceil$ rounds only elements $i \in [\![K]\!]$ satisfying: $\Lambda_i \leq \Lambda_{(N_\epsilon)}$ are eliminated. Therefore, if $1/\epsilon > \Lambda^*$, we have : $\Lambda_{(N_\epsilon)} = \Lambda^*$ and all the remaining experts are optimal (i.e. in $\mathcal{S}^*$). Hence the mean of any two experts in $\mathcal{S}$ satisfies: $R(\hat{g}) \leq R^*$.

Now suppose that $1/\epsilon < \Lambda^*$. We have for the last $T - \lceil \tau_{N_\epsilon} \rceil$ rounds all the experts in $\mathcal{S}_\epsilon^c$ were eliminated (hence there was at most $|\mathcal{S}_\epsilon|$ non-eliminated experts). Let $(\hat{k}, \hat{l})$ denote the pair output by algorithm M-2 after T rounds, we have:

$$\begin{aligned}
T_{\hat{k}\hat{l}}(T) &\geq \log(C_\epsilon \delta^{-1}) \frac{T - \tau_{N_\epsilon}}{|\mathcal{S}_\epsilon|^2} \\
&\geq 289 \frac{\log(C_\epsilon \delta^{-1})}{\epsilon} \\
&\geq c \log(KT\delta^{-1}) \frac{1}{\epsilon},
\end{aligned}$$

where c is a numerical constant, we used (15) for the second line, and a simple calculation to obtain the last line. Using Lemma 6, we obtain the desired conclusion.

6 Proof of Theorem M-4.1

In this section we will show that for C large enough, if $\mathcal{A}$ holds, we have:

$$R(\hat{g}) - R^* \lesssim \epsilon. \tag{16}$$

Let i^* be an arbitrary element of $\mathcal{S}^*$. Denote T_i the number of queries required to eliminate an expert $i \in [\![K]\!]$. T_i is upper bounded by the number of data points needed to have: $\Delta_{i^*i} > 0$, which would lead to the elimination of i by i^*. The following claim, which is a consequence of Lemma 7, provides this upper bound.

Claim 3. *If $\mathcal{A}$ holds, let $i \in [\![K]\!]$ be a suboptimal expert ($\Lambda_i < +\infty$). We have:*

$$T_i \leq 289 \log\big(KC\delta^{-1}\big)\Lambda_i.$$

Proof. Lemma 5 shows that experts $i^* \in \mathcal{S}^*$ are never eliminated if $\mathcal{A}$ is true. Using Lemma 7, the number of queries required for the elimination of a suboptimal expert i by expert i^*, satisfies:

$$T_i \leq 289 \log\big(KC\delta^{-1}\big)\Lambda_i.$$

□

Let $\epsilon \geq 0$. Recall that $\mathcal{S}_\epsilon$ is defined by:

$$\mathcal{S}_\epsilon := \left\{ i \in [\![K]\!] : \Lambda_i > \frac{1}{\epsilon} \right\}$$

Suppose that we have:

$$C > 578 \left(\sum_{i \in \mathcal{S}_\epsilon^c} \Lambda_i + |\mathcal{S}_\epsilon| \min\left\{ \frac{1}{\epsilon}; \Lambda^* \right\} \right) \log\left(K\delta^{-1} \left(\sum_{i \in \mathcal{S}_\epsilon^c} \Lambda_i + |\mathcal{S}_\epsilon| \min\left\{ \frac{1}{\epsilon}; \Lambda^* \right\} \right) \right),$$

We therefore have using Lemma 2:

$$C > 289 \log\big(KC\delta^{-1}\big) \left(\sum_{i \in \mathcal{S}_\epsilon^c} \Lambda_i + |\mathcal{S}_\epsilon| \min\left\{ \frac{1}{\epsilon}; \Lambda^* \right\} \right).$$

Let us denote by C_1 the total number of queries received by all the experts in $\mathcal{S}_\epsilon$ and by C_2 the total number of queries received by the remaining experts. We therefore have: $C = C_1 + C_2$. In order to show that at a certain round, all the experts in $\mathcal{S}_\epsilon^c$ were eliminated, it suffices to prove that:

$$C_1 \geq |\mathcal{S}_\epsilon| \max_{i \in \mathcal{S}_\epsilon^c} T_i,$$

since the inequality above shows that the budget is not totally consumed after round $\max_{i \in \mathcal{S}_\epsilon^c} T_i$ where all elements in $\mathcal{S}_\epsilon^c$ where eliminated.

Claim 3 provides the following upper bound for C_2:

$$C_2 \leq 289 \log\big(KC\delta^{-1}\big) \sum_{i \in \mathcal{S}_\epsilon^c} \Lambda_i.$$

We therefore have:

$$\begin{aligned}
C_1 &= C - C_2 \\
&\geq 289 \log\big(KC\delta^{-1}\big) \left(\sum_{i \in \mathcal{S}_\epsilon^c} \Lambda_i + |\mathcal{S}_\epsilon| \min\left\{ \frac{1}{\epsilon}; \Lambda^* \right\} \right) - C_2 \\
&\geq 289 \log\big(KC\delta^{-1}\big) \left(\sum_{i \in \mathcal{S}_\epsilon^c} \Lambda_i + |\mathcal{S}_\epsilon| \min\left\{ \frac{1}{\epsilon}; \Lambda^* \right\} \right) - 289 \log\big(KC\delta^{-1}\big) \sum_{i \in \mathcal{S}_\epsilon^c} \Lambda_i.
\end{aligned}$$

Hence:

$$C_1 \geq 289 \log\big(KC\delta^{-1}\big)|\mathcal{S}_\epsilon| \min\left\{\frac{1}{\epsilon}; \Lambda^*\right\} \tag{17}$$

Recall that by definition of $\mathcal{S}_\epsilon$, using Claim 3 we have:

$$\max_{i \in \mathcal{S}_\epsilon^c} T_i \leq 289 \log\big(KC\delta^{-1}\big) \min\left\{\frac{1}{\epsilon}; \Lambda^*\right\},$$

hence:

$$C_1 \geq |\mathcal{S}_\epsilon| \max_{i \in \mathcal{S}_\epsilon^c} T_i.$$

This shows that $S \subseteq \mathcal{S}_\epsilon$. We have two possibilities: if $\frac{1}{\epsilon} < \Lambda^*$, the selected pair $(F_{\bar{k}}, F_{\bar{l}}) \in S \times S$ satisfies:

$$T_{\bar{k}\bar{l}} = \min\{T_{\bar{k}}, T_{\bar{l}}\} \geq \frac{C_1}{|\mathcal{S}_\epsilon|}.$$

Using (17), we have:

$$T_{\bar{k}\bar{l}} \geq 289 \log\big(KC\delta^{-1}\big)\frac{1}{\epsilon}. \tag{18}$$

Observe that Lemma 6 applies in this setting. In particular, the total number of rounds T of algorithm M-1, satisfy: $T \leq C$. Hence, it holds

$$R\left(\frac{F_{\hat{k}} + F_{\hat{l}}}{2}\right) - R^* \leq c\, B \frac{\log(KC\delta^{-1})}{T_{\bar{k}\bar{l}}}.$$

We conclude by injecting inequality (18) in the bound above. We therefore have:

$$R(\hat{g}) - R^* \leq cB\, \epsilon,$$

where c is an absolute constant.

If $\frac{1}{\epsilon} > \Lambda^*$, by definition of Λ^* and the fact that $\mathcal{S} \subseteq \mathcal{S}_\epsilon$, we conclude that only the optimal experts (i.e. the experts i such that $R_i = R^*$) remain when the budget is totally consumed. Hence combining any 2 of these expert will lead to the bound: $R(\hat{g}) \leq R^*$.

7 Proof of lower bounds

The lemma below gives a lower bound for the problem of estimating the parameter describing a Bernoulli random variable.

Lemma 8 ([1], Lemma 5.1)**.** Suppose that α is a random variable uniformly distributed on $\{\alpha_-, \alpha_+\}$, where $\alpha_- = 1/2 - \epsilon/2$ and $\alpha_+ = 1/2 + \epsilon/2$, with $0 < \epsilon < 1$. Suppose that $\xi_1, \ldots, \xi_m$ are i.i.d $\{0,1\}$-valued random variables with $\mathbb{P}(\xi_i = 1) = \alpha$ for all i. Let f be a function from $\{0,1\} \to \{\alpha_-, \alpha_+\}$. Then it holds:

$$\mathbb{P}(f(\xi_1, \ldots, \xi_m) \neq \alpha) > \frac{1}{4}\left(1 - \sqrt{1 - \exp\left(\frac{-2\lceil m/2 \rceil \epsilon^2}{1 - \epsilon^2}\right)}\right).$$

7.1 Proof of Lemma M-6.1

Let $T > 0$ and consider an convex combination of experts $\hat{g}$ output after full observation of T training rounds. We will construct two experts F_1 and F_2 and a target variable Y and we will show that, for these variables, a strategy for our problem ($m = 2$ and $p = 1$) gives a solution to the problem in Lemma 8. Finally we will use the lower bound from this lemma.

For $\theta \in [0,1]$, let $\mathbb{P}_\theta$ denote the probability distribution of T i.i.d. draws $Y_1, \ldots, Y_T$ of Bernoulli variables or parameter θ, while $F_{1,t} = 0$ and $F_{2,t} = 1$ almost surely for $t \in [\![T]\!]$. Let α be a variable that is uniformly distributed on $\{\alpha_-, \alpha_+\}$ with $\alpha_\pm = \frac{1}{2} \pm \frac{\epsilon}{2}$, and $\epsilon \in (0,1)$ is a parameter to be tuned subsequently; let the training obervations be drawn according to $\mathbb{P}_\alpha$. Since $p = 1$, the output $\hat{g}$ is either F_1 or F_2. Define $f : \{0,1\}^T \to \{\alpha_-, \alpha_+\}$ such that given $(Y_1, \ldots, Y_T)$, f outputs $\frac{1}{2} - \frac{\epsilon}{2}$ if $\hat{g} = F_1$ and $\frac{1}{2} + \frac{\epsilon}{2}$ if $\hat{g} = F_2$. By construction we have that the events $\{f = \alpha\}$ and

$\{R(\hat{g}) = \min\{R_1, R_2\}\}$ are equivalent. Using Lemma 8 and setting $\epsilon = \frac{c_0}{\sqrt{T}}$ where c_0 is a constant such that the lower bound in Lemma 8 is equal to 0.1, we have:

$$\mathbb{P}\left(R(\hat{g}) - \min\{R_1, R_2\} \geq \frac{c_0}{\sqrt{T}}\right) > 0.1.$$

Due to the randomization of α, the above probability is the average of the corresponding event under $\mathbb{P}_{\alpha_-}$ and $\mathbb{P}_{\alpha_+}$. Therefore, under at least one of these two training distributions, the deviation event has a probability at least 0.05.

7.2 Proof of Lemma M-6.2

The gist of the proof is the following. We will construct a distribution with two experts that are very correlated. In this situation, going from a weighted average of the two experts to a single expert with the largest weight does not change the prediction risk much, and so we could find a single expert with small risk if the weighted average has small risk. On the other hand, since the agent only observes one expert per training round, from their point of view the observational distribution is identical as if the experts were independent – the correlation cannot be observed. Therefore the same strategy could be used to find the best expert in the independent case. This contradicts the lower bounds in this case (which is a standard bandit setting), therefore it is impossible to pick consistently a weighted average with small risk in a situation where the correlations cannot be observed.

Let $T > 0$ be fixed. We consider the particular setting where the target variable Y is identically 0, and the expert predictions F_1 and F_2 are two (non independent) Bernoulli random variables. We define a distribution $\mathbb{P}_-$ for (F_1, F_2) such that:

- the marginal distribution of F_1 is Bernoulli of parameter $\alpha_- = \frac{1}{2} - \frac{\epsilon}{2}$;
- the marginal distribution of F_2 is Bernoulli of parameter $\alpha_+ = \frac{1}{2} + \frac{\epsilon}{2}$;
- it holds that $\mathbb{P}_-(F_1 F_2 = 1) = \alpha_-$.

Note that this can be easily constructed as $F_1 = \mathbf{1}\{U \leq \alpha_-\}; F_2 = \mathbf{1}\{U \leq \alpha_+\}$, where U is a uniform variable on $[0, 1]$. Let $\mathbb{P}_+$ be defined similarly with the role of F_1 and F_2 reversed. Here, ϵ is a positive parameter to be tuned later. We denote R_-, R_+ for the prediction risks under distributions $\mathbb{P}_-, \mathbb{P}_+$. We have $R_-(F_1) = R_+(F_2) = \alpha_-$, $R_-(F_2) = R_+(F_1) = \alpha_+$, and $R^* = \alpha_-$ is the same under $\mathbb{P}_-$ and $\mathbb{P}_+$.

Let us be given an arbitrary training observation strategy π (prescribing at each training round which expert to observe based only on past observations), and output a convex combination of experts $\hat{g}$. This output is a convex combination of F_1 and F_2, hence it is characterized by the weight of F_1, which we denote $\hat{\alpha}$. The parameter $\hat{\alpha}$ depends on the observed data. We also define $\hat{f}$ associated to this training strategy, that outputs F_1 if $\hat{\alpha} > \frac{1}{2}$ and F_2 otherwise. Finally, let us denote $\mathbb{Q}_\pi^+$ the distribution of the training data observed by the agent when the T experts opinions are drawn i.i.d. from $\mathbb{P}_-$ and the agent observes the expert advices following strategy π; and define $\mathbb{Q}_\pi^-$ similarly.

Define the event $\mathcal{A}_+ := \left\{R_+(\hat{g}) - R^* \geq \frac{1}{4}\epsilon\right\}$ and similarly $\mathcal{A}_-$. In the remainder of the proof, we will show, using Bretagnolle-Hubert inequality (Theorem 14.2 in [2]), that either $\mathbb{Q}_\pi^-(\mathcal{A}_-)$ or $\mathbb{Q}_\pi^+(\mathcal{A}_+)$ is lower bounded by a positive constant.

We have under the distribution $\mathbb{P}_-$:

$$\begin{aligned}
R_-(\hat{g}) - R_-(\hat{f}) &= \mathbb{E}_-\left[(\hat{\alpha}F_1 + (1-\hat{\alpha})F_2)^2\right] - \mathbb{E}_-\left[\left(\mathbb{1}\left(\hat{\alpha} > \frac{1}{2}\right)F_1 + \mathbb{1}\left(\hat{\alpha} \leq \frac{1}{2}\right)F_2\right)^2\right] \\
&= \epsilon(1-\hat{\alpha})^2 - \epsilon\left(1 - \mathbb{1}\left(\hat{\alpha} > \frac{1}{2}\right)\right) \\
&\geq -\frac{3}{4}\epsilon.
\end{aligned}$$

Note that the above estimate crucially depends on the fact that F_1, F_2 are not independent under $\mathbb{P}_-$. In view of the above, the event $\mathcal{A}_-$ is implied by $R_-(\hat{f}) - R^* = \epsilon$. Similarly, $\mathcal{A}_+$ is implied by

$R_+(\hat{f}) - R^* = \epsilon$. Hence:

$$\mathbb{Q}_\pi^-(\mathcal{A}_-) + \mathbb{Q}_\pi^+(\mathcal{A}_+) \geq \mathbb{Q}_\pi^-\Big(R_-(\hat{f}) - R^* = \epsilon\Big) + \mathbb{Q}_\pi^+\Big(R_+(\hat{f}) - R^* = \epsilon\Big)$$
$$= \mathbb{Q}_\pi^-\Big(\hat{f} = F_2\Big) + \mathbb{Q}_\pi^+\Big(\hat{f} \neq F_2\Big).$$

Now we use Bretagnolle-Hubert inequality:

$$\mathbb{Q}_\pi^-(f = F_2) + \mathbb{Q}_\pi^+(f \neq F_2) \geq \frac{1}{2}\exp\big(-D\big(\mathbb{Q}_\pi^-, \mathbb{Q}_\pi^+\big)\big),$$

where $D(\mathbb{Q}_\pi^-, \mathbb{Q}_\pi^+)$ is the relative entropy between $\mathbb{Q}_\pi^-$ and $\mathbb{Q}_\pi^+$. In order to conclude, we need an upper bound on $D(\mathbb{Q}_\pi^-, \mathbb{Q}_\pi^+)$. Since the agent only observes one expert in each round according to strategy π, the distribution of the observed data $\mathbb{Q}_\pi^-$ or $\mathbb{Q}_\pi^+$ is unchanged if we replace the generating distributions $\mathbb{P}_-$ or $\mathbb{P}_+$ by distributions having the same marginals, but for which F_1 and F_2 are independent. Therefore, the observational distributions $\mathbb{Q}_\pi^-, \mathbb{Q}_\pi^+$ are equivalent to that of the observational distributions, under the same strategy, of a canonical bandit model with two arms. We can then use the divergence decomposition formula (Lemma 15.1 of [2]) to upper bound $D(\mathbb{Q}_\pi^-, \mathbb{Q}_\pi^+)$; denoting $\mathbb{P}_-^{(1)}$, $\mathbb{P}_-^{(2)}$ the marginals of $\mathbb{P}_-$ and similarly for $\mathbb{P}_+$, it holds

$$D\big(\mathbb{Q}_\pi^-, \mathbb{Q}_\pi^+\big) = \mathbb{E}_-[T_1]D(\mathbb{P}_-^{(1)}, \mathbb{P}_+^{(1)}) + \mathbb{E}_-[T_2]D(\mathbb{P}_-^{(2)}, \mathbb{P}_+^{(2)}),$$

where the expectation $\mathbb{E}_-[.]$ is with respect to the probability distribution $\mathbb{Q}_\pi^-$ and T_i denotes the total number of rounds where the advice of expert F_i was queried using the strategy π. We have: $T_1 + T_2 = T$ almost surely, and $D(\mathbb{P}_-^{(1)}, \mathbb{P}_+^{(1)}) = D(\mathbb{P}_-^{(2)}, \mathbb{P}_+^{(2)}) \leq 4\epsilon^2$ provided $\epsilon \leq \frac{1}{2}$. Therefore:

$$\mathbb{Q}_\pi^-(\mathcal{A}_-) + \mathbb{Q}_\pi^+(\mathcal{A}_+) \geq \frac{1}{2}\exp\big(-4\epsilon^2 T\big).$$

This shows that there exists a probability distribution $\mathbb{P} \in \{\mathbb{P}_-, \mathbb{P}_+\}$ for the experts advices and the target variable such that the prediction $\hat{g}$ satisfies:

$$\mathbb{P}(R(\hat{g}) - R^* \geq \epsilon) \geq \exp\big(-4\epsilon^2 T\big),$$

We conclude by choosing $\epsilon = \frac{1}{2\sqrt{T}}$.

8 Intermediate case: $m \geq 3, p = 2$

In this section we assume that the learner is allowed to access more than two experts advices per round. We show that this leads to an improvement of the bound in Theorem M-5.2. We consider the following extension of Algorithm M-2:

Algorithm 1 Intermediate case

Input m, L and ρ.
Initialization: $S \leftarrow [\![K]\!]$.
for $T = 1, 2, \ldots$ **do**
  Sample a subset $\mathcal{M}$ of size m from $[\![K]\!]$ uniformly at random.
  Query the advice of experts in $\mathcal{M}$ and update the corresponding quantities.
  For all i, j: If $\Delta'_{ij} > 0$: $S \leftarrow S \setminus \{j\}$.
end for
On interrupt: Let $\hat{k} \in S$ and let $\hat{l} \leftarrow \underset{j \in S}{\operatorname{argmax}}\ \hat{d}_{\hat{k}j}$.
Return $\frac{1}{2}\big(F_{\hat{k}} + F_{\hat{l}}\big)$.

Theorem 9. (Instance independent bound) Suppose Assumption M-1 holds. Let $T \geq 1$, and denote $\hat{g}$ the output of Algorithm 1 with inputs (m, L, ρ) in round T. If $m \geq 3$, then with probability at least $1 - \delta$:

$$R(\hat{g}) \leq \min_{i \in [\![K]\!]} R_i + cB\,\frac{(K/m)^2 \log\big(2TK\delta^{-1}\big)}{T},$$

where c is an absolute constant.

Proof. Let $i, j \in [\![K]\!]$, denote $T_{ij}(T)$ the total number of rounds where the advice of expert i and j were jointly queried. We have: $T_{ij}(T) = \sum_{t=1}^{T} \mathbb{1}\{i \text{ and } j \text{ were jointly queried at round } t\}$. We conclude that $T_{ij}(T)$ is the sum of T independent and identically distributed Bernoulli variables with parameter: $\frac{m(m-1)}{K(K-1)}$. We therefore have the following consequence of Bernstein concentration inequality, with probability at least $1-\delta$, for all $i, j \in [\![K]\!]$ and $T \geq K$:

$$|T_{ij}(T) - \mathbb{E}[T_{ij}(T)]| \leq \sqrt{2T\frac{m(m-1)}{K(K-1)}\log(2KT/\delta)} + \frac{1}{3}\log(2KT/\delta). \tag{19}$$

Suppose that δ satisfies:

$$\log(2KT/\delta) \leq \frac{1}{16}\frac{m^2}{K^2}T.$$

Then we have:

$$\sqrt{2T\frac{m(m-1)}{K(K-1)}\log(2KT/\delta)} + \frac{1}{3}\log(2KT/\delta) \leq \frac{1}{2}\frac{m(m-1)}{K(K-1)}T, \tag{20}$$

Observe that the result of Lemma 6 still holds in this setting for non-eliminated elements (experts in S_T), since the elimination criterion for an expert j, which consists of the existence of i such that $\Delta'_{ij} > 0$, is the same as in Algorithm M-2. Let $\hat{g}$ denote the output of Algorithm 1, we conclude that if $\mathcal{A}$ and (19) hold for all i, j and T, we have:

$$R(\hat{g}) - R_{i^*} \leq \kappa\frac{\log\big(KT\delta^{-1}\big)}{T_{\hat{k}\hat{l}}(T)}, \tag{21}$$

where κ is a constant depending only η, L and ρ. Finally, we use (20). We therefore have with probability at least $1-4\delta$:

$$R(\hat{g}) \leq \min_{i\in[\![K]\!]} R_i + c\,B\frac{(K/m)^2\log\big(2TK\delta^{-1}\big)}{T}.$$

Now suppose that δ satisfies:

$$\log(2KT/\delta) \geq \frac{1}{16}\frac{m^2}{K^2}T,$$

then it holds:

$$\frac{(K/m)^2\log\big(2TK\delta^{-1}\big)}{T} \geq \frac{1}{16}.$$

We conclude that for $\bar{c} = \max\{c, 16\}$ we have:

$$R(\hat{g}) - \min_{i\in[\![K]\!]} R_i \leq B \leq \bar{c}B\frac{(K/m)^2\log\big(2TK\delta^{-1}\big)}{T}.$$

□

References

[1] M. Anthony and P. L. Bartlett. *Neural network learning: Theoretical foundations*. Cambridge University Press, 2009.

[2] T. Lattimore and C. Szepesvári. *Bandit algorithms*. Cambridge University Press, 2020.

[3] A. Maurer and M. Pontil. Empirical Bernstein bounds and sample-variance penalization. In *COLT 2009 - The 22nd Conference on Learning Theory, Montreal, Quebec, Canada, June 18-21, 2009*, 2009.