Minimizing Regret with Label Efficient Prediction

Nicolò Cesa-Bianchi, Gábor Lugosi, *Member, IEEE,* and Gilles Stoltz

***Abstract*— We investigate label efficient prediction, a variant, proposed by Helmbold and Panizza, of the problem of prediction with expert advice. In this variant the forecaster, after guessing the next element of the sequence to be predicted, does not observe its true value unless he asks for it, which he cannot do too often. We determine matching upper and lower bounds for the best possible excess prediction error, with respect to the best possible constant predictor, when the number of allowed queries is fixed. We also prove that Hannan consistency, a fundamental property in game-theoretic prediction models, can be achieved by a forecaster issuing a number of queries growing to infinity at a rate just slightly faster than logarithmic in the number of prediction rounds.**

***Index Terms*— label efficient prediction, prediction with expert advice, individual sequences, on-line learning**

I. Introduction

Prediction with expert advice, a framework introduced about fifteen years ago in learning theory, may be viewed as a direct generalization of the theory of repeated games, a field pioneered by Blackwell and Hannan in the mid-fifties. At a certain level of abstraction, the common subject of these studies is the problem of forecasting each element y_t of an unknown "target" sequence given the knowledge of the previous elements $y_1, \ldots, y_{t-1}$. The forecaster's goal is to predict the target sequence almost as well as any forecaster forced to use the same guess all the times. We call this the sequential prediction problem. To provide a suitable parameterization of the problem, we assume that the set from which the forecaster picks its guesses is finite, of size $N > 1$, while the set to which the target sequence elements belong may be of arbitrary cardinality. A real-valued bounded loss function ℓ is then used to quantify the discrepancy between each outcome y_t and the forecaster's guess for y_t. The pioneering results of Hannan's [1] and Blackwell [2] showed that randomized forecasters exist whose excess cumulative loss (or regret), with respect to the loss of any constant forecaster, grows sub-linearly in the length n of the target sequence, and this holds for any individual target sequence. In particular, both Blackwell and Hannan found the optimal growth rate, $\Theta(\sqrt{n})$, of the regret as a function of the sequence length n when no assumption other than boundedness is made on the loss ℓ. Only relatively recently, Cesa-Bianchi, Freund, Haussler, Helmbold, Schapire, and Warmuth [3] have revealed that the correct dependence on N in the minimax regret rate is $\Theta(\sqrt{n \ln N})$.

Game theorists, information theorists, and learning theorists, who independently studied the sequential prediction model, addressed the fundamental question of whether a sub-linear regret rate is achievable in case the past outcomes $y_1, \ldots, y_{t-1}$ are not entirely accessible when computing the guess for y_t. In this work we investigate a variant of sequential prediction known as *label efficient prediction*. In this model, originally proposed by Helmbold and Panizza [4], after choosing its guess at time t the forecaster decides whether to query the outcome y_t. However, the forecaster is limited in the number $\mu(n)$ of queries he can issue within a given time horizon n. In the case $n \to \infty$, we prove that Hannan consistency (i.e., regret growing sub-linearly with probability one) can be achieved under the only condition $\mu(n)/(\log(n) \log\log(n)) \to \infty$. Moreover, in the finite-horizon case, we show that any forecaster issuing at most $m = \mu(n)$ queries must suffer a regret of at least order $n\sqrt{(\ln N)/m}$ on some outcome sequence of length n, and we show a randomized forecaster achieving this regret to within constant factors.

The problem of label efficient prediction is closely related to other frameworks in which the forecaster has a limited access to the outcomes. Examples include prediction under partial monitoring (see, e.g., Mertens, Sorin, and Zamir [5], Rustichini [6], Piccolboni, and Schindelhauer [7], Mannor and Shimkin [8], Cesa-Bianchi, Lugosi, and Stoltz [9]), the multi-armed bandit problem (see Baños [10], Megiddo [11], Foster and Vohra [12], Hart and Mas Colell [13], Auer, Cesa-Bianchi, Freund, and Schapire [14], and Auer [15]), and the "apple tasting" problem proposed by Helmbold, Littlestone, and Long [16].

II. Sequential prediction and the label efficient model

The sequential prediction problem is parameterized by a number $N > 1$ of player actions, by a set $\mathcal{Y}$ of outcomes, and by a loss function ℓ. The loss function has domain $\{1, \ldots, N\} \times \mathcal{Y}$ and takes values in a bounded real interval, say $[0, 1]$. Given an unknown mechanism generating a sequence $y_1, y_2, \ldots$ of elements from $\mathcal{Y}$, a prediction strategy, or forecaster, chooses an action $I_t \in \{1, \ldots, N\}$ incurring a loss $\ell(I_t, y_t)$. A crucial assumption in this model is that the forecaster can choose I_t only based on information related to the past outcomes $y_1, \ldots, y_{t-1}$. That is, the forecaster's decision must not depend on any of the future outcomes. In the label efficient model, after choosing I_t the forecaster decides whether to issue a query to access y_t. If no query is

N. Cesa-Bianchi is with the Dipartimento di Scienze dell'Informazione, Università di Milano, via Comelico 39, 20135 Milano, Italy, `cesa-bianchi@dsi.unimi.it`, G. Lugosi is with the Department of Economics, Universitat Pompeu Fabra, Ramon Trias Fargas 25-27, 08005 Barcelona, Spain, `lugosi@upf.es`, G. Stoltz is with the Département de Mathématiques et Applications, Ecole Normale Supérieure, 45 rue d'Ulm, 75005 Paris, France, `gilles.stoltz@ens.fr`.

The work of the first two authors was supported in part by the IST Programme of the European Community under the PASCAL Network of Excellence IST-2002-506778. The work of the second author was supported by the Spanish Ministry of Science and Technology and FEDER, grant BMF2003-03324. An extended abstract appeared in the *Proceedings of the 17th Annual Conference on Learning Theory*, Springer, 2004.

LABEL EFFICIENT PREDICTION

Parameters: number N of actions, outcome space $\mathcal{Y}$, loss function ℓ, query rate $\mu : \mathbb{N} \to \mathbb{N}$.

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

(1) the environment chooses the next outcome $y_t \in \mathcal{Y}$ without revealing it;
(2) the forecaster chooses an action $I_t \in \{1, \ldots, N\}$;
(3) each action i incurs loss $\ell(i, y_t)$;
(4) if less than $\mu(t)$ queries have been issued so far, the forecaster may issue a new query to obtain the outcome y_t; if no query is issued then y_t remains unknown.

Fig. 1. Label efficient prediction as a game between the forecaster and the environment.

issued, then y_t remains unknown. In other words, I_t does not depend on all the past outcomes $y_1, \ldots, y_{t-1}$, but only on the queried ones. The label efficient model is best described as a repeated game between the forecaster, choosing actions, and the environment, choosing outcomes (see Figure 1).

The cumulative loss of the forecaster on a sequence $y_1, y_2, \ldots$ of outcomes is denoted by

$$\widehat{L}_n = \sum_{t=1}^{n} \ell(I_t, y_t) \qquad \text{for } n \geq 1.$$

As the forecasting strategies we consider may be randomized, each I_t is viewed as a random variable. All probabilities and expectations are understood with respect to the σ-algebra of events generated by the sequence of random choices of the forecaster. We compare the forecaster's cumulative loss $\widehat{L}_n$ with those of the N constant forecasters $L_{i,n} = \ell(i, y_1) + \ldots + \ell(i, y_n)$, $i = 1, \ldots, N$.

In this paper we devise label efficient forecasting strategies whose expected regret

$$\mathbb{E}\left[\widehat{L}_n - \min_{i=1,\ldots,N} L_{i,n}\right]$$

grows sub-linearly in n for any sequence $y_1, y_2, \ldots$ of outcomes, that is, for any strategy of the environment whenever $\mu(n) \to \infty$. Note that the quantities $L_{1,n}, \ldots, L_{N,n}$ are random. Indeed, as argued in Section III, in general the outcomes y_t may depend on the forecaster's past random choices. Via a more refined analysis, we also prove the stronger result

$$\widehat{L}_n - \min_{i=1,\ldots,N} L_{i,n} = o(n) \qquad \text{a.s.} \tag{1}$$

for any sequence $y_1, y_2, \ldots$ of outcomes and whenever $\mu(n)/(\log(n) \log\log(n)) \to \infty$. The almost sure convergence is with respect to the auxiliary randomization the forecaster has access to. Property (1), known as *Hannan consistency* in game theory, rules out the possibility that the regret is much larger than its expected value with a significant probability.

Parameters: Real numbers $\eta > 0$ and $0 \leq \varepsilon \leq 1$.
Initialization: $\boldsymbol{w}_1 = (1, \ldots, 1)$.

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

(1) draw an action I_t from $\{1, \ldots, N\}$ according to the distribution
$$p_{i,t} = \frac{w_{i,t}}{\sum_{j=1}^{N} w_{j,t}} , \qquad i = 1, \ldots, N ;$$
(2) draw a Bernoulli random variable Z_t such that $\mathbb{P}[Z_t = 1] = \varepsilon$;
(3) if $Z_t = 1$ then obtain y_t and compute
$$w_{i,t+1} = w_{i,t}\, e^{-\eta\, \ell(i,y_t)/\varepsilon} \qquad \text{for each } i = 1, \ldots, N$$
else, let $\boldsymbol{w}_{t+1} = \boldsymbol{w}_t$.

Fig. 2. The label efficient exponentially weighted average forecaster.

III. A LABEL EFFICIENT FORECASTER

We start by considering the finite-horizon case in which the forecaster's goal is to control the regret after n predictions, where n is fixed in advance. In this restricted setup we also assume that at most $m = \mu(n)$ queries can be issued, where μ is the query rate function. However, we do not impose any further restriction on the distribution of these m queries in the n time steps, that is, $\mu(t) = m$ for $t = 1, \ldots, n$. We introduce a simple forecaster whose expected regret is bounded by $n\sqrt{2(\ln N)/m}$.

It is easy to see that in order to achieve a nontrivial performance, a forecaster must use randomization in determining whether a label should be revealed or not. It turns out that a simple biased coin is sufficient for our purpose. The strategy we propose, sketched in Figure 2, uses an i.i.d. sequence $Z_1, Z_2, \ldots, Z_n$ of Bernoulli random variables such that $\mathbb{P}[Z_t = 1] = 1 - \mathbb{P}[Z_t = 0] = \varepsilon$ and asks the label y_t to be revealed whenever $Z_t = 1$. Here $\varepsilon > 0$ is a parameter of the strategy. (Typically, we take $\varepsilon \approx m/n$ so that the number of solicited labels during n rounds is about m. Note that this way the forecaster may ask the value of more than m labels, but we ignore this detail as it can be dealt with by a simple adjustment.) Our label efficient forecaster uses the *estimated losses*

$$\widetilde{\ell}(i, y_t) \stackrel{\text{def}}{=} \begin{cases} \ell(i, y_t)/\varepsilon & \text{if } Z_t = 1, \\ 0 & \text{otherwise.} \end{cases}$$

Let $\boldsymbol{p}_t = (p_{1,t}, \ldots, p_{N,t})$ and let v_1^t denote the prefix $(v_1, \ldots, v_t)$ of an arbitrary sequence $(v_1, v_2, \ldots)$. Then

$$\mathbb{E}[\,\widetilde{\ell}(i, y_t) \mid Z_1^{t-1}, I_1^{t-1}] = \ell(i, y_t) , \tag{2}$$

$$\begin{aligned} &\mathbb{E}[\,\widetilde{\ell}(\boldsymbol{p}_t, y_t) \mid Z_1^{t-1}, I_1^{t-1}] \\ &\quad = \ell(\boldsymbol{p}_t, y_t) = \mathbb{E}[\,\ell(I_t, y_t) \mid Z_1^{t-1}, I_1^{t-1}] , \end{aligned} \tag{3}$$

hold for each t, where

$$\ell(\boldsymbol{p}_t, y_t) = \sum_{i=1}^{N} p_{i,t}\, \ell(i, y_t) \quad \text{and} \quad \widetilde{\ell}(\boldsymbol{p}_t, y_t) = \sum_{i=1}^{N} p_{i,t}\, \widetilde{\ell}(i, y_t) .$$

Note that the conditioning on Z_1^{t-1} and I_1^{t-1} is necessary because of the two following reasons: first, $\boldsymbol{p}_t$ depends both on the past realizations of the random choices of the forecaster Z_1^{t-1} (see the third step in the algorithm of Figure 2) and on the past outcomes y_1^{t-1}; second, y_t is a function of both Z_1^{t-1} and I_1^{t-1}, as the environment is allowed to determine y_t after playing the game up to time $t-1$ (see Figure 1). For technical reasons, we sometimes consider a weaker model (which we call the *oblivious adversary*) where the sequence $y_1, y_2, \ldots$ of outcomes chosen by the environment is deterministic and independent of the forecaster random choices. This is equivalent to a game in which the environment must fix the sequence of outcomes before the game begins. The oblivious adversary model is reasonable in some scenarios, in which the forecaster's predictions have no influence on the environment. Clearly, any result proven in the standard model also holds in the oblivious adversary model.

The quantities $\widetilde{\ell}(i, y_t)$ may be considered as unbiased estimates of the true losses $\ell(i, y_t)$. The label efficient forecaster of Figure 2 is an exponentially weighted average forecaster using such estimates instead of the observed losses. The expected performance of this strategy may be bounded as follows.

Theorem 1: Fix a time horizon n and consider the label efficient forecaster of Figure 2 run with parameters $\varepsilon = m/n$ and $\eta = (\sqrt{2m \ln N})/n$. Then, the expected number of revealed labels equals m and

$$\mathbb{E}\left[\widehat{L}_n - \min_{i=1,\ldots,N} L_{i,n}\right] \le n\sqrt{\frac{2\ln N}{m}} \ .$$

In the sequel, for each $i = 1, \ldots, N$, we write

$$\widetilde{L}_{i,n} = \sum_{t=1}^{n} \widetilde{\ell}(i, y_t) \ .$$

PROOF. The proof is a simple adaptation of [17, Theorem 3.1]. The starting point is the following inequality (see also [7, Theorem 1]):

$$\sum_{t=1}^{n} \widetilde{\ell}(\boldsymbol{p}_t, y_t) - \min_{i=1,\ldots,N} \widetilde{L}_{i,n} \le \frac{\ln N}{\eta} + \frac{\eta}{2}\sum_{t=1}^{n}\sum_{j=1}^{N} \widetilde{\ell}(j, y_t)^2 p_{j,t} \ .$$

Since $\widetilde{\ell}(j, y_t) \in [0, 1/\varepsilon]$ for all j and y_t, the second term on the right-hand side may be bounded by $(\eta/(2\varepsilon)) \sum_{t=1}^{n}\sum_{j=1}^{N} \widetilde{\ell}(j, y_t) p_{j,t}$ and therefore we get, for all n,

$$\sum_{t=1}^{n} \widetilde{\ell}(\boldsymbol{p}_t, y_t)\left(1 - \frac{\eta}{2\varepsilon}\right) \le \widetilde{L}_{i,n} + \frac{\ln N}{\eta} \qquad i = 1, \ldots, N \ . \tag{4}$$

Taking expectations on both sides and substituting the values of η and ε yields the desired result. ∎

Remark 1.1: In the oblivious adversary model, Theorem 1 (and similarly Theorems 2 and 10 below) can be strengthened as follows. Consider the "lazy" forecaster of Figure 3 that keeps on choosing the same action as long as no new queries are issued. For this forecaster Theorems 1 and 2 hold with the additional statement that, with probability 1, the number of changes of an action, that is the number of steps where $I_t \neq I_{t+1}$, is at most the number of queried labels (by construction of the lazy forecaster). To prove the regret bound, note that we derive the statement of Theorem 1 by taking averages on both sides of (4), and then applying (2) and (3). Note that (4) holds for *every* realization of the random variables $I_1, \ldots, I_n$ and $Z_1, \ldots, Z_n$. Therefore, as the lazy forecaster differs from the forecaster of Figure 2 only in the distribution of $I_1, \ldots, I_n$, inequality (4) holds for the lazy forecaster as well. In the oblivious adversary model y_t does not depend on $I_1, \ldots, I_{t-1}$; thus, by construction, $\boldsymbol{p}_t$ does not depend on $I_1, \ldots, I_{t-1}$ either. Therefore, we can take averages with respect to $I_1, \ldots, I_{t-1}$ obtaining the following version of (3) for the lazy forecaster,

$$\mathbb{E}\left[\widetilde{\ell}(\boldsymbol{p}_t, y_t) \mid Z_1^{t-1}\right] = \sum_{i=1}^{N} \ell(i, y_t)\, p_{i,t} = \mathbb{E}\left[\ell(I_t, y_t) \mid Z_1^{t-1}\right] \ .$$

Since (2) holds as well when the conditioning is limited to $Z_1, \ldots, Z_{t-1}$, we can derive for the lazy forecaster the same bounds as in Theorem 1 (and Theorem 2). Note also that the result holds even when y_t is allowed to depend on $Z_1, \ldots, Z_{t-1}$.

Parameters: Real numbers $\eta > 0$ and $0 \le \varepsilon \le 1$.
Initialization: $\boldsymbol{w}_1 = (1, \ldots, 1)$, $Z_0 = 1$.

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

(1) if $Z_{t-1} = 1$ then draw an action I_t from $\{1, \ldots, N\}$ according to the distribution

$$p_{i,t} = \frac{w_{i,t}}{\sum_{j=1}^{N} w_{j,t}} \ , \qquad i = 1, \ldots, N \ ;$$

otherwise, let $I_t = I_{t-1}$;

(2) draw a Bernoulli random variable Z_t such that $\mathbb{P}[Z_t = 1] = \varepsilon$;

(3) if $Z_t = 1$ then obtain y_t and compute

$$w_{i,t+1} = w_{i,t}\, e^{-\eta\, \ell(i,y_t)/\varepsilon} \qquad \text{for each } i = 1, \ldots, N$$

else, let $\boldsymbol{w}_{t+1} = \boldsymbol{w}_t$.

Fig. 3. The lazy label efficient exponentially weighted average forecaster for the oblivious adversary model.

A. *Bounding the regret with high probability*

Theorem 1 guarantees that the expected per-round regret converges to zero whenever $m \to \infty$ as $n \to \infty$. The next result shows that this regret is, with overwhelming probability, bounded by a quantity proportional to $n\sqrt{(\ln N)/m}$.

Theorem 2: Fix a time horizon n and a number $\delta \in (0, 1)$. Consider the label efficient forecaster of Figure 2 run with parameters

$$\varepsilon = \max\left\{0, \frac{m - \sqrt{2m \ln(4/\delta)}}{n}\right\} \quad \text{and} \quad \eta = \sqrt{\frac{2\varepsilon \ln N}{n}} \ .$$

Then, with probability at least $1-\delta$, the number of revealed labels is at most m and for all $t=1,\dots,n$,

$$\widehat{L}_t - \min_{i=1,\dots,N} L_{i,t} \le 2n\sqrt{\frac{\ln N}{m}} + 6n\sqrt{\frac{\ln(4N/\delta)}{m}} \ .$$

Before proving Theorem 2, note that if $\delta \le 4Ne^{-m/8}$, then the right-hand side of the inequality is greater than n and therefore the statement is trivial. Thus, we may assume throughout the proof that $\delta > 4Ne^{-m/8}$. This ensures that

$$\varepsilon \ge m/(2n) > 0 \ . \tag{5}$$

We need a number of preliminary lemmas. The first is obtained by a simple application of Bernstein's inequality (see Lemma 15).

Lemma 3: The probability that the strategy asks for more than m labels is at most $\delta/4$.

PROOF. Note that the number $M = \sum_{t=1}^n Z_t$ of labels asked by the algorithm is binomially distributed with parameters n and ε and therefore, writing $\gamma = m/n - \varepsilon = n^{-1}\sqrt{2m\ln(4/\delta)}$, it satisfies

$$\begin{aligned}\mathbb{P}[M>m] &= \mathbb{P}[M-\mathbb{E}M > n\gamma] \le e^{-n\gamma^2/(2\varepsilon+2\gamma/3)} \\ &\le e^{-n^2\gamma^2/2m} \le \frac{\delta}{4}\end{aligned}$$

where we used Bernstein's inequality (see Lemma 15) in the second step and the definition of γ in the last two steps. ∎

Lemma 4: With probability at least $1-\delta/4$, for all $t=1,\dots,n$,

$$\sum_{s=1}^t \ell(\boldsymbol{p}_s, y_s) \le \sum_{s=1}^t \widetilde{\ell}(\boldsymbol{p}_s, y_s) + \frac{4}{\sqrt{3}}n\sqrt{\frac{\ln(4/\delta)}{m}} \ .$$

Furthermore, with probability at least $1-\delta/4$, for all $i=1,\dots,N$ and for all $t=1,\dots,n$,

$$\widetilde{L}_{i,t} \le L_{i,t} + \frac{4}{\sqrt{3}}n\sqrt{\frac{\ln(4N/\delta)}{m}} \ .$$

PROOF. The proofs of both inequalities rely on the same techniques, namely the application of Bernstein's inequality for martingales combined with Doob's maximal inequality. We therefore focus on the first one, and indicate the modifications needed for the second one.

We introduce the sequence $X_s = \ell(\boldsymbol{p}_s, y_s) - \widetilde{\ell}(\boldsymbol{p}_s, y_s)$, $s=1,\dots,n$, which is a martingale difference sequence with respect to the filtration generated by the (Z_s, I_s), $s=1,\dots,n$. Defining $u = (4/\sqrt{3})n\sqrt{(1/m)\ln(4/\delta)}$ and the martingale $M_t = X_1 + \ldots + X_t$, our goal is to show that

$$\mathbb{P}\left[\max_{t=1,\dots,n} M_t > u\right] \le \frac{\delta}{4} \ .$$

For all $s=1,\dots,n$, we note that

$$\begin{aligned}&\mathbb{E}\left[X_s^2 \mid Z_1^{s-1}, I_1^{s-1}\right] \\ &\quad= \mathbb{E}\left[\left(\ell(\boldsymbol{p}_s,y_s) - \widetilde{\ell}(\boldsymbol{p}_s,y_s)\right)^2 \mid Z_1^{s-1}, I_1^{s-1}\right] \\ &\quad\le \mathbb{E}\left[\widetilde{\ell}(\boldsymbol{p}_s,y_s)^2 \mid Z_1^{s-1}, I_1^{s-1}\right] \le 1/\varepsilon \ ,\end{aligned}$$

so that summing over s, we have $V_t \le n/\varepsilon$ for all $t=1,\dots,n$.

We now apply Lemma 15 with $x=u$, $v=n/\varepsilon$, and $K=1/\varepsilon$ (since $|X_s| \le 1/\varepsilon$ with probability 1 for all s). This yields

$$\begin{aligned}\mathbb{P}\left[\max_{t=1,\dots,n} M_t > x\right] &= \mathbb{P}\left[\max_{t=1,\dots,n} M_t > u \text{ and } V_n \le \frac{n}{\varepsilon}\right] \\ &\le \exp\left(-\frac{u^2}{2\left(n/\varepsilon + u/(3\,\varepsilon)\right)}\right) \ .\end{aligned}$$

Using $\ln(4/\delta) \le m/8$ implied by the assumption $\delta > 4Ne^{-m/8}$, we see that $u \le n$, which, combined with (5), shows that

$$\frac{u^2}{2\left(n/\varepsilon + u/(3\,\varepsilon)\right)} \ge \frac{u^2}{(8/3)\,n/\varepsilon} \ge \frac{3u^2\,m}{16\,n^2} = \ln\frac{\delta}{4}$$

and this proves the first inequality.

To prove the second inequality note that, by the arguments above, for each fixed i we have

$$\mathbb{P}\left[\forall\, t \le n \quad \widetilde{L}_{i,t} > L_{i,t} + (4/\sqrt{3})\,n\sqrt{\frac{\ln(4N/\delta)}{m}}\right] \le \frac{\delta}{4N} \ .$$

The proof is concluded by a union-of-events bound. ∎

PROOF OF THEOREM 2. When $m \le \ln N$, the bound given by the theorem is trivial, so we only need to consider the case when $m \ge \ln N$. Then (5) implies that $1-\eta/(2\varepsilon) \ge 0$. Thus, a straightforward combination of Lemmas 3 and 4 with (4) shows that, with probability at least $1-3\delta/4$, the strategy asks for at most m labels and for all $t=1,\dots,n$,

$$\begin{aligned}&\sum_{s=1}^t \ell(\boldsymbol{p}_s,y_s)\left(1-\frac{\eta}{2\varepsilon}\right) \\ &\quad\le \min_{i=1,\dots,N} L_{i,t} + \frac{8}{\sqrt{3}}n\sqrt{\frac{1}{m}\ln\frac{4N}{\delta}} + \frac{\ln N}{\eta} \ ,\end{aligned}$$

which, since $\sum_{s=1}^t \ell(\boldsymbol{p}_s,y_s) \le n$ for all $t \le n$, implies, for all $t=1,\dots,n$,

$$\begin{aligned}&\sum_{s=1}^t \ell(\boldsymbol{p}_s,y_s) - \min_{i=1,\dots,N} L_{i,t} \\ &\quad\le \frac{n\eta}{2\varepsilon} + \frac{8}{\sqrt{3}}n\sqrt{\frac{1}{m}\ln\frac{4N}{\delta}} + \frac{\ln N}{\eta} \\ &\quad= 2n\sqrt{\frac{\ln N}{m}} + \frac{8}{\sqrt{3}}n\sqrt{\frac{1}{m}\ln\frac{4N}{\delta}}\end{aligned}$$

by our choice of η and using $1/(2\varepsilon) \le n/m$ derived from (5). The proof is finished by noting that the Hoeffding-Azuma inequality (for maximal processes, see [18]) implies that, with probability at least $1-\delta/4$, for all $t=1,\dots,n$,

$$\begin{aligned}\widehat{L}_t &= \sum_{s=1}^t \ell(I_s,y_s) \le \sum_{s=1}^t \ell(\boldsymbol{p}_s,y_s) + \sqrt{\frac{n}{2}\ln\frac{4}{\delta}} \\ &\le \sum_{s=1}^t \ell(\boldsymbol{p}_s,y_s) + n\sqrt{\frac{1}{2\,m}\ln\frac{4N}{\delta}}\end{aligned}$$

since $m \le n$. ∎

B. Hannan consistency

Theorem 1 does not directly imply Hannan consistency of the associated forecasting strategy because the regret bound does not hold uniformly over the sequence length n. However, using standard dynamical tuning techniques (such as the "doubling trick" described in [3]) Hannan consistency can be achieved. The main quantity that arises in the analysis is the query rate $\mu(n)$, that is the number of queries that can be issued up to time n. The next result shows that Hannan consistency is achievable whenever $\mu(n)/(\log(n)\log\log(n)) \to \infty$.

Corollary 5: Let $\mu : \mathbb{N} \to \mathbb{N}$ be any nondecreasing integer-valued function such that

$$\lim_{n\to\infty} \frac{\mu(n)}{\log_2(n)\log_2\log_2(n)} = \infty \; .$$

Then there exists a Hannan consistent randomized label efficient forecaster that issues at most $\mu(n)$ queries in the first n predictions, for any $n \in \mathbb{N}$.

PROOF. The algorithm we consider divides time into consecutive epochs of increasing lengths $n_r = 2^r$ for $r = 0, 1, 2, \ldots$. In the r-th epoch (of length 2^r) the algorithm runs the forecaster of Theorem 2 with parameters $n = 2^r$, $m = m_r$, and $\delta_r = 1/(1+r)^2$, where m_r will be determined by the analysis (without loss of generality, we assume the forecaster always asks at most m_r labels in each epoch r). Our choice of δ_r and the Borel-Cantelli lemma implies that the bound of Theorem 2 holds for all but finitely many epochs. Denote the (random) index of the last epoch in which the bound does not hold by $\hat{R}$. Let $L^{(r)}$ be cumulative loss of the best action in epoch r and let $\widehat{L}^{(r)}$ be the cumulative loss of the forecaster in the same epoch. Introduce $R(n) = \lfloor \log_2 n \rfloor$. Then, by Theorem 2 and by definition of $\hat{R}$, for each n and for each realization of I_1^n and Z_1^n we have

$$\begin{aligned}
\widehat{L}_n - L_n^* &\le \sum_{r=0}^{R(n)-1} \left(\widehat{L}^{(r)} - L^{(r)} \right) + \sum_{t=2^{R(n)}}^{n} \ell(I_t, y_t) \\
&\quad - \sum_{t=2^{R(n)}}^{n} \min_{j=1,\ldots,N} \ell(j, y_t) \\
&\le \sum_{r=0}^{\hat{R}} 2^r + 8 \sum_{r=\hat{R}+1}^{R(n)} 2^r \sqrt{\frac{\ln(4N(r+1)^2)}{m_r}} \; .
\end{aligned}$$

This, the finiteness of $\hat{R}$, and $1/n \le 2^{-R(n)}$, imply that with probability 1,

$$\limsup_{n\to\infty} \frac{\widehat{L}_n - L_n^*}{n} \le 8 \limsup_{R\to\infty} 2^{-R} \sum_{r=0}^{R} 2^r \sqrt{\frac{\ln(4N(r+1)^2)}{m_r}} \; .$$

Cesaro's lemma ensures that the $\limsup$ above equals zero as soon as $m_r/\ln r \to +\infty$. It remains to see that the latter condition is satisfied under the additional requirement that the forecaster does not issue more than $\mu(n)$ queries up to time n. This is guaranteed whenever $m_0 + m_1 + \ldots + m_{R(n)} \le \mu(n)$ for each n. Denote by ϕ the largest nondecreasing function such that

$$\phi(t) \le \frac{\mu(t)}{(1+\log_2 t)\log_2(1+\log_2 t)} \qquad \text{for all } t = 1, 2, \ldots$$

Parameters: Real number $0 \le \varepsilon \le 1$.

Initialization: $t = 1$.

For each epoch $r = 0, 1, 2, \ldots$,

1) let $K_r = 4^r (2 \ln N)/\varepsilon$;
2) initialize $\widetilde{L}_i(r) = 0$ for all $i = 1, \ldots, N$;
3) restart the forecaster of Figure 2 choosing ε and $\eta_r = \sqrt{(2\varepsilon \ln N)/K_r}$;
4) **while** $\min_i \widetilde{L}_i(r) \le K_r - 1/\varepsilon$ **do**:
   (a) denote by I_t the action chosen by the forecaster of Figure 2, and let $Z_t = 1$ if it asks for the label y_t, $Z_t = 0$ otherwise;
   (b) if $Z_t = 1$, then obtain the outcome y_t and update the estimated losses, for all $i = 1, \ldots, N$, as
   $$\widetilde{L}_i(r) := \widetilde{L}_i(r) + \ell(i, y_t)/\varepsilon \; ;$$
   (c) $t := t + 1$.

Fig. 4. A doubling version of the label efficient exponentially weighted average forecaster.

As μ grows faster than $\log_2(n)\log_2\log_2(n)$, we have that $\phi(t) \to +\infty$. Thus, choosing $m_0 = 0$, and $m_r = \lfloor \phi(2^r)\log_2(1+r) \rfloor$, we indeed ensure that $m_r/\ln r \to +\infty$. Furthermore, using that m_r is nondecreasing as a function of r, and using the monotonicity of ϕ,

$$\begin{aligned}
\sum_{r=0}^{R(n)} m_r &\le (R(n)+1)\phi(2^{R(n)})\log_2(1+R(n)) \\
&\le (1+\log_2 n)\phi(n)\log_2(1+\log_2 n) \le \mu(n)
\end{aligned}$$

and this concludes the proof. ∎

IV. Improvements for small losses

We now prove a refined bound in which the factors $n\sqrt{(\ln N)/m}$ of Theorem 2 are replaced by quantities of the order of $\sqrt{nL_n^*(\ln N)/m + (n/m)\ln N}$ in case of an oblivious adversary, and $\sqrt{nL_n^*(\ln(Nn))/m + (n/m)\ln(Nn)}$ in case of a non-oblivious one, where L_n^* is the cumulative loss of the best action,

$$L_n^* = L_n^*(y_1^n) = \min_{i=1,\ldots,N} \sum_{t=1}^{n} \ell(i, y_t) \; .$$

In particular, we recover the behavior already observed by Helmbold and Panizza [4] for oblivious adversaries in the case $L_n^* = 0$.

This is done by introducing a modified version of the forecaster of Figure 2, which performs a doubling trick over the estimated losses $\widetilde{L}_{i,t}$, $t = 1, \ldots, n$ (see Figure 4), and whose performance is studied below through several applications of Bernstein's lemma.

Similarly to [17, Section 4], we propose in Figure 4 a forecaster which uses a doubling trick based on the estimated losses of each action $i = 1, \ldots, N$. We denote the estimated accumulated loss of this algorithm by

$$\widetilde{L}_{A,n} = \sum_{t=1}^{n} \widetilde{\ell}(\boldsymbol{p}_t, y_t)$$

and prove the following inequality.

Lemma 6: For any $0 \le \varepsilon \le 1$, the forecaster of Figure 4 achieves, for all $n = 1, 2, \ldots$,

$$\widetilde{L}_{A,n} \le \widetilde{L}_n^* + 8\sqrt{2}\sqrt{\left(\widetilde{L}_n^* + 1/\varepsilon\right)\frac{\ln N}{\varepsilon}} + \frac{4 \ln N}{\varepsilon}$$

where

$$\widetilde{L}_n^* = \min_{i=1,\ldots,N} \widetilde{L}_{i,n} \ .$$

PROOF. The proof is divided in three steps. We first deal with each epoch, then sum the estimated losses over the epochs, and finally bound the total number R of different epochs (i.e., the final value of r). Let S_r and T_r be the first and last time steps completed on epoch r (where for convenience we define $T_R = n$). Thus, epoch r consists of trials $S_r, S_r + 1, \ldots, T_r$. We denote the estimated cumulative loss of the forecaster at epoch r by

$$\widetilde{L}_A(r) = \sum_{t=S_r}^{T_r} \widetilde{\ell}(\boldsymbol{p}_t, y_t)$$

and the estimated cumulative losses of the actions $i = 1, \ldots, N$ at epoch r by

$$\widetilde{L}_i(r) = \sum_{t=S_r}^{T_r} \widetilde{\ell}(i, y_t) \ .$$

Inequality (4) ensures that for epoch r, and for all $i = 1, \ldots, N$,

$$\left(1 - \frac{\eta_r}{2\varepsilon}\right)\widetilde{L}_A(r) \le \widetilde{L}_i(r) + \frac{\ln N}{\eta_r}$$

so dividing both terms by the quantity $1 - \eta_r/(2\varepsilon)$ (which is more than $1/2$ due to the choice of K_r), we get

$$\widetilde{L}_A(r) \le \widetilde{L}_i(r) + \frac{\eta_r}{\varepsilon}\widetilde{L}_i(r) + 2\frac{\ln N}{\eta_r} \ .$$

The stopping condition now guarantees that $\min_i \widetilde{L}_i(r) \le K_r$, hence, substituting the value of η_r, we have proved that for epoch r,

$$\widetilde{L}_A(r) \le \min_{i=1,\ldots,N} \widetilde{L}_i(r) + 2\sqrt{2}\sqrt{\frac{K_r \ln N}{\varepsilon}} \ .$$

Summing over $r = 0, \ldots, R$, we get

$$\begin{aligned}
\widetilde{L}_{A,n} &\le \sum_{r=0}^{R} \min_{i=1,\ldots,N} \widetilde{L}_i(r) + \sum_{r=0}^{R} 2\sqrt{2}\sqrt{\frac{K_r \ln N}{\varepsilon}} \\
&\le \min_{i=1,\ldots,N} \widetilde{L}_{i,n} + 2\sqrt{2}\sqrt{\frac{K_0 \ln N}{\varepsilon}}\left(2^{R+1} - 1\right) \quad (6)
\end{aligned}$$

It remains to bound the number R of epochs, or alternatively, to bound $2^{R+1} - 1$. Assume first that $R \ge 1$. In particular,

$$\begin{aligned}
\widetilde{L}_n^* &= \min_{i=1,\ldots,N} \widetilde{L}_{i,n} \ge \min_{i=1,\ldots,N} \widetilde{L}_i(R-1) \\
&> K_{R-1} - 1/\varepsilon = 4^{R-1}K_0 - 1/\varepsilon
\end{aligned}$$

so

$$2^{R-1} \le \sqrt{\left(\widetilde{L}_n^* + 1/\varepsilon\right)\frac{1}{K_0}} \ .$$

The above is implied by

$$2^{R+1} - 1 \le 1 + 4\sqrt{\left(\widetilde{L}_n^* + 1/\varepsilon\right)\frac{1}{K_0}}$$

which also holds for $R = 0$. Applying the last inequality to (6) concludes the proof. ∎

We now state and prove a bound that holds in the most general (non-oblivious) adversarial model.

Theorem 7: The label efficient forecaster of Figure 4, run with

$$\varepsilon = \frac{m - \sqrt{2m \ln(4/\delta)}}{n}$$

ensures that, with probability $1 - \delta$, the algorithm does not ask for more than m labels and for all $t = 1, \ldots, n$,

$$\widehat{L}_t - L_t^* \le U(L_n^*) + \sqrt{2\left(1 + L_n^* + U(L_n^*)\right)\ln\frac{4n}{\delta}} + \frac{1}{2}\ln\frac{4n}{\delta}$$

where

$$\begin{aligned}
&U(L_n^*) \\
&= 20\sqrt{\frac{n}{m}L_n^* \ln\frac{4Nn}{\delta}} + 32\left(\frac{n}{m}\ln\frac{4Nn}{\delta}\right)^{3/4}(L_n^*)^{1/4} \\
&\quad +10\left(\frac{n}{m}\ln\frac{4Nn}{\delta}\right)^{7/8}(L_n^*)^{1/8} + 75\frac{n}{m}\ln\frac{4Nn}{\delta} \\
&\le 137 \times \max\left\{\sqrt{\frac{n}{m}L_n^* \ln\frac{4Nn}{\delta}} \ , \frac{n}{m}\ln\frac{4Nn}{\delta}\right\} \ .
\end{aligned}$$

We remark here that the bound of the theorem is an improvement over that of Theorem 2 as soon as L_n^* grows slower than $n/\sqrt{\ln n}$. (For $L_n^* \sim n$ however, these bounds are worse, at least in the case of non-oblivious adversary, see Theorem 10 below for a refined bound for the case of an oblivious adversary.)

First we relate $\widetilde{L}_n^*$ to L_n^*, and $\widetilde{L}_{A,n}$ to $\bar{L}_{A,n}$, where

$$\bar{L}_{A,n} = \sum_{t=1}^{n} \ell(\boldsymbol{p}_t, y_t)$$

is the sum of the conditional expectations of the instantaneous losses, and then substitute the obtained inequalities in the bound of Lemma 6.

Lemma 8: With probability $1 - \delta/2$, the following $2n$ inequalities hold simultaneously: for all $t = 1, \ldots, n$,

$$\begin{aligned}
\widetilde{L}_t^* &\le L_t^* + 2\sqrt{\frac{n}{m}L_n^* \ln\frac{4Nn}{\delta}} + 4\frac{n}{m}\ln\frac{4Nn}{\delta} \ , \\
\widetilde{L}_{A,t} &\ge \bar{L}_{A,t} - \left(2\sqrt{\frac{n}{m}\bar{L}_{A,n}\ln\frac{4n}{\delta}} + 4\frac{n}{m}\ln\frac{4n}{\delta}\right) \ .
\end{aligned}$$

PROOF. We prove that each of both lines holds with probability at least $1 - \delta/4$. As the proofs are similar, we concentrate on the first one only. For all $i = 1, \ldots, N$, we apply Corollary 16 with $X_t = \widetilde{\ell}(i, y_t) - \ell(i, y_t)$, $t = 1, \ldots, n$, which forms a martingale difference sequence (with respect to the filtration generated by (I_t, Z_t), $t = 1, \ldots, n$). With the notation of the corollary, $K = 1/\varepsilon$, and V_n is smaller than $L_{i,n}/\varepsilon$, which shows that (for a given i), with probability at least $1-\delta/(4N)$,

$$\max_{t=1,\ldots,n} \left(\widetilde{L}_{i,t} - L_{i,t}\right) \le \sqrt{2\left(\frac{1}{\varepsilon^2} + \frac{L_{i,n}}{\varepsilon}\right)\ln\frac{4Nn}{\delta}} + \frac{\sqrt{2}}{3\varepsilon}\ln\frac{4Nn}{\delta} \ .$$

The proof is concluded by using $\sqrt{x+y} \le \sqrt{x} + \sqrt{y}$ for $x, y \ge 0$, $1/\varepsilon \le 2n/m$ (derived from (5)), $\ln(4Nn/\delta) \ge 1$ and the union-of-events bound. ∎

Lemma 9: With probability at least $1-\delta/2$,

$$\forall\, t = 1,\ldots,n \qquad \bar{L}_{A,t} - L_t^* \le U(L_n^*) \ ,$$

where $U(L_n^*)$ is as in Theorem 7.

PROOF. We combine the inequalities of Lemma 8 with Lemma 6, and perform some trivial upper bounding, to get that, with probability $1-\delta/2$, for all $t = 1,\ldots,n$,

$$\bar{L}_{A,t} \le L_t^* + 2\sqrt{\frac{n}{m}\bar{L}_{A,n}\ln\frac{4Nn}{\delta}} + 18\sqrt{\frac{n}{m}L_n^*\ln\frac{4Nn}{\delta}} + 23\,(L_n^*)^{1/4}\left(\frac{n}{m}\ln\frac{4Nn}{\delta}\right)^{3/4} + 56\frac{n}{m}\ln\frac{4Nn}{\delta} \ .$$

An application of Lemma 19 concludes the proof. ∎

PROOF OF THEOREM 7. Lemma 3 shows that with probability at least $1-\delta/4$, the number of queried labels is less than m. Using the notation of Corollary 16, we consider the martingale difference sequence formed by $X_t = \ell(I_t, y_t) - \ell(\boldsymbol{p}_t, y_t)$, with associated sum of conditional variances $V_n \le \bar{L}_{A,n}$ and increments bounded by 1. Corollary 16 then shows that with probability $1-\delta/4$,

$$\max_{t=1,\ldots,n}\left(\widehat{L}_t - \bar{L}_{A,t}\right) \le \sqrt{2\left(1+\bar{L}_{A,n}\right)\ln\frac{4n}{\delta}} + \frac{\sqrt{2}}{3}\ln\frac{4n}{\delta} \ .$$

We conclude the proof by applying Lemma 9 and a union-of-events bound. ∎

In the oblivious adversary model, the bound of Theorem 7 can be strengthened as follows.

Theorem 10: In the oblivious adversary model, the label efficient forecaster of Figure 4, run with

$$\varepsilon = \frac{m - \sqrt{2m\ln(4/\delta)}}{n}$$

ensures that with probability $1-\delta$, the algorithm does not ask for more than m labels and that

$$\forall\, t = 1,\ldots,n \qquad \widehat{L}_t - L_t^* \le B(L_n^*) + 2\sqrt{(L_n^* + B(L_n^*))\ln\frac{4}{\delta}}$$

where

$$\begin{aligned} B(L_n^*) &= 21\sqrt{\frac{n}{m}L_n^*\ln\frac{4N}{\delta}} + 39\left(\frac{n}{m}\ln\frac{4N}{\delta}\right)^{3/4}(L_n^*)^{1/4} \\ &\quad + 15\left(\frac{n}{m}\ln\frac{4N}{\delta}\right)^{7/8}(L_n^*)^{1/8} + 59\frac{n}{m}\ln\frac{4N}{\delta} \\ &\le 134\max\left(\sqrt{\frac{n}{m}L_n^*\ln\frac{4N}{\delta}}\ , \frac{n}{m}\ln\frac{4N}{\delta}\right) \ . \end{aligned}$$

Observe that the order of magnitude of the bound of Theorem 10 is always at least as good as that of Theorem 2 and is better as soon as L_n^* grows slower than n.

The proof of Theorem 10 is based on combining Lemma 6 with two applications of Bernstein's inequality, but here, one of these applications is a backward call to Bernstein's inequality: usually, one can handle the predictable quadratic variation of the studied martingale, and Bernstein's inequality is then a useful concentration result for the martingale. In the case of the second step below we know the deviations of the martingale (formed by $\widetilde{L}_{A,n}$), but we are interested in the behavior of its predictable quadratic variation (equal to $\bar{L}_{A,n}$). The two quantities are related by a "backwards" use of Bernstein's lemma.

First step: Relating estimated losses to the cumulative loss of the best action: We relate $\widetilde{L}_n^*$ and $\widetilde{L}_{A,n}$ to L_n^* by using Bernstein's inequality (Lemma 15). First we point out the difference between oblivious and non-oblivious adversaries. More precisely, to apply Lemma 15 rather than Corollary 16, we need upper bounds K_i for all $L_{i,n} = L_{i,n}(y_1^n)$ (we exceptionally make the dependence on the played outcomes explicit) which are independent of I_1^n and Z_1^n. In case of an oblivious adversaries, the outcome sequence y_1^n is chosen in advance, and $K_i = L_{i,n}(y_1^n)$ is a suitable choice. This is not the case for non-oblivious adversaries whose behavior may take the actions of the forecaster into account (see the previous section).

Observe the similarity of the first statement of the following lemma to Lemmas 4 and 8.

Lemma 11: When facing an oblivious adversary, with probability $1-\delta/4$,

$$\forall\, t = 1,\ldots,n, \qquad \widetilde{L}_t^* \le L_t^* + 2\sqrt{\frac{n}{m}L_n^*\ln\frac{4N}{\delta}} + \frac{n}{m}\ln\frac{4N}{\delta} \ .$$

Consequently, with probability $1-\delta/4$,

$$\forall\, t = 1,\ldots,n, \qquad \widetilde{L}_{A,t} \le L_t^* + A(L_n^*) \ , \tag{7}$$

where

$$\begin{aligned} A(L_n^*) &= 18\sqrt{\frac{n}{m}L_n^*\ln\frac{4N}{\delta}} \\ &\quad + 23\left(\frac{n}{m}\ln\frac{4N}{\delta}\right)^{3/4}(L_n^*)^{1/4} + 37\frac{n}{m}\ln\frac{4N}{\delta} \ . \end{aligned}$$

PROOF. For all $i = 1,\ldots,N$, we may apply Lemma 15 with $X_t = \widetilde{\ell}(i, y_t) - \ell(i, y_t)$, $t = 1,\ldots,n$, which forms a martingale difference sequence with respect to the filtration generated by Z_t, $t = 1,\ldots,n$. With the notation of Lemma 15, $V_n \le L_{i,n}/\varepsilon \le 2n\,L_{i,n}/m$, which is indeed independent of the Z_t, and simple algebra and the union-of-events bound conclude the proof of the first statement. The second one follows from a combination of the first one with Lemma 6. ∎

Second step: Bernstein's inequality used backwards: Next we relate $\bar{L}_{A,n}$ to $\widetilde{L}_{A,n}$ (and thus to L_n^*, via Lemma 11). This is done by using Bernstein's lemma (Lemma 15) once again, but backwards.

Lemma 12: For oblivious adversaries, with probability at least $1-\delta/2$,

$$\forall t=1,\ldots,n \qquad \bar{L}_{A,t} - L_t^* \le B(L_n^*) \ ,$$

where $B(L_n^*)$ is as in Theorem 10.

PROOF. Consider $A(L_n^*)$ as in Lemma 11 and fix a real number $x_0 > A(L_n^*)$. Recall the function ϕ_K defined in the statement of Lemma 15. Then (7) and the union-of-events bound imply that, for $\lambda > 0$ such that $\lambda - \phi_1(\lambda)/\varepsilon > 0$,

$$\begin{aligned}
&\mathbb{P}\left[\max_{t=1,\ldots,n}\left(\bar{L}_{A,t}-L_t^*\right) > x_0\right] \\
&\le \frac{\delta}{4} + \mathbb{P}\Bigg[\max_{t=1,\ldots,n}\left(\bar{L}_{A,t}-L_t^*\right) > x_0 \\
&\qquad\qquad \text{and } \max_{t=1,\ldots,n}\left(\widetilde{L}_{A,t}-L_t^*\right) \le A(L_n^*)\Bigg] \\
&\le \frac{\delta}{4} \\
&\quad + \mathbb{P}\Bigg[\max_{t=1,\ldots,n}\exp\Bigg(\left(\lambda - \frac{\phi_1(\lambda)}{\varepsilon}\right) \\
&\qquad \cdot\left(\bar{L}_{A,t}-L_t^*\right) - \lambda\left(\widetilde{L}_{A,t}-L_t^*\right)\Bigg) \\
&\qquad > \exp\left(\left(\lambda - \frac{\phi_1(\lambda)}{\varepsilon}\right)x_0 - \lambda A(L_n^*)\right)\Bigg] \\
&\le \frac{\delta}{4} + \mathbb{P}\Bigg[\max_{t=1,\ldots,n}\exp\left(\lambda\left(\bar{L}_{A,t}-\widetilde{L}_{A,t}\right) - \frac{\phi_1(\lambda)}{\varepsilon}\bar{L}_{A,t}\right) \\
&\qquad > \exp\left(\left(\lambda - \frac{\phi_1(\lambda)}{\varepsilon}\right)x_0 - \lambda A(L_n^*) - \frac{\phi_1(\lambda)}{\varepsilon}L_n^*\right)\Bigg] \qquad (8)
\end{aligned}$$

We introduce the martingale difference sequence (with increments bounded by 1) $X_t = \ell(\boldsymbol{p}_t, y_t) - \widetilde{\ell}(\boldsymbol{p}_t, y_t)$. The conditional variances satisfy

$$\mathbb{E}\left[X_t^2 \mid Z_1^{t-1}\right] \le \mathbb{E}\left[\widetilde{\ell}(\boldsymbol{p}_t, y_t)^2 \mid Z_1^{t-1}\right] \le \frac{\ell(\boldsymbol{p}_t, y_t)}{\varepsilon}$$

so that, using the notation of Lemma 15, $V_n \le \bar{L}_{A,n}/\varepsilon$.

By Lemma 15, $\exp\big(\lambda\big(\bar{L}_{A,t} - \widetilde{L}_{A,t}\big) - \phi_1(\lambda)V_t\big)$ for $t = 1, 2, \ldots$ is a nonnegative supermartingale. Hence, using Doob's maximal inequality, we get

$$\begin{aligned}
&\mathbb{P}\Bigg[\max_{t=1,\ldots,n}\exp\left(\lambda\left(\bar{L}_{A,t}-\widetilde{L}_{A,n}\right) - \frac{\phi_1(\lambda)}{\varepsilon}\bar{L}_{A,t}\right) \\
&\qquad > \exp\left(\left(\lambda - \frac{\phi_1(\lambda)}{\varepsilon}\right)x_0 - \lambda A(L_n^*) - \frac{\phi_1(\lambda)}{\varepsilon}L_n^*\right)\Bigg] \\
&\le \mathbb{P}\Bigg[\max_{t=1,\ldots,n}\exp\left(\lambda\left(\bar{L}_{A,t}-\widetilde{L}_{A,t}\right) - \phi_1(\lambda)V_t\right) \\
&\qquad > \exp\left(\lambda\left(x_0 - A(L_n^*)\right) - \frac{\phi_1(\lambda)}{\varepsilon}\left(x_0 + L_n^*\right)\right)\Bigg] \\
&\le \exp\left(\lambda\left(A(L_n^*) - x_0\right) + \frac{\phi_1(\lambda)}{\varepsilon}\left(x_0 + L_n^*\right)\right) \ . \qquad (9)
\end{aligned}$$

Now, choose

$$\lambda = \frac{x_0 - A(L_n^*)}{2\left(x_0 + L_n^*\right)}\,\varepsilon \ .$$

$\lambda \le \varepsilon/2 \le 1$, and therefore, using $\phi_1(t) \le t^2$ for $t \le 1$, we have proved that $\lambda - \phi_1(\lambda)/\varepsilon > 0$. Thus, (8) and (9) imply

$$\begin{aligned}
&\mathbb{P}\left[\max_{t=1,\ldots,n}\left(\bar{L}_{A,t}-L_t^*\right) > x_0\right] \\
&\le \frac{\delta}{4} + \exp\left(\lambda\left(A(L_n^*) - x_0\right) + \frac{\lambda^2}{\varepsilon}\left(x_0 + L_n^*\right)\right) \\
&= \frac{\delta}{4} + \exp\left(-\frac{\left(A(L_n^*) - x_0\right)^2}{4\left(x_0 + L_n^*\right)}\,\varepsilon\right) \ .
\end{aligned}$$

It suffices to find a $x_0 > A(L_n^*)$ such that

$$\frac{\left(A(L_n^*) - x_0\right)^2}{4\left(x_0 + L_n^*\right)}\,\varepsilon = \ln\frac{\delta}{4} \ .$$

One such choice is

$$x_0 = A(L_n^*) + \frac{2\ln\frac{\delta}{4}}{\varepsilon} + 2\sqrt{\frac{\ln\frac{\delta}{4}}{\varepsilon}}\sqrt{L_n^* + A(L_n^*) + \frac{\ln\frac{\delta}{4}}{\varepsilon}} \ .$$

Substituting the value of $A(L_n^*)$ yields the statement of the lemma. ∎

Third step: Conclusion of the proof of Theorem 10: Lemma 3 shows that, with probability at least $1-\delta/4$, the number of queried labels is less than m. We then consider the martingale difference sequence formed by $X_t = \ell(I_t, y_t) - \ell(\boldsymbol{p}_t, y_t)$, with associated sum of conditional variances $V_n \le \bar{L}_{A,n}$ and increments bounded by 1. Lemma 15 yields

$$\begin{aligned}
&\mathbb{P}\left[\max_{t=1,\ldots,n}\left(\widehat{L}_t - \bar{L}_{A,t}\right) > u \text{ and } \bar{L}_{A,n} \le L_n^* + B(L_n^*)\right] \\
&\qquad\qquad \le \exp\left(-\frac{u^2}{4\left(L_n^* + B(L_n^*)\right)}\right)
\end{aligned}$$

provided that $u \le 3(L_n^* + B(L_n^*))$. Lemma 12 together with a union-of-events bound and the choice

$$u = 2\sqrt{\left(L_n^* + B(L_n^*)\right)\ln\frac{4}{\delta}}$$

concludes the proof.

V. A LOWER BOUND FOR LABEL EFFICIENT PREDICTION

Here we show that the performance bounds proved in Section III for the label efficient exponentially weighted average forecaster are essentially unimprovable in the strong sense that no other label efficient forecasting strategy can have a significantly better performance for all problems. Denote the set of natural numbers by $\mathbb{N} = \{1, 2, \ldots\}$.

Theorem 13: There exist an outcome space $\mathcal{Y}$, a loss function $\ell : \mathbb{N} \times \mathcal{Y} \to [0,1]$, and a universal constant $c > 0$ such that, for all $N \ge 2$ and for all $n \ge m \ge 20\frac{e}{1+e}\ln(N-1)$, the cumulative (expected) loss of any (randomized) forecaster that uses actions in $\{1, \ldots, N\}$ and asks for at most m labels while predicting a sequence of n outcomes satisfies the inequality

$$\begin{aligned}
&\sup_{y_1,\ldots,y_n \in \mathcal{Y}}\left(\mathbb{E}\left[\sum_{t=1}^n \ell(I_t, y_t)\right] - \min_{i=1,\ldots,N}\sum_{t=1}^n \ell(i, y_t)\right) \\
&\qquad\qquad \ge c\,n\sqrt{\frac{\ln(N-1)}{m}} \ .
\end{aligned}$$

In particular, we prove the theorem for $c = \dfrac{\sqrt{e}}{(1+e)\sqrt{5(1+e)}}$.

PROOF. First, we define $\mathcal{Y} = [0,1]$ and ℓ. Given $y \in [0,1]$, we denote by $(y_1, y_2, \ldots)$ its dyadic expansion, that is, the unique sequence not ending with infinitely many zeros such that

$$y = \sum_{k \ge 1} y_k \, 2^{-k} \ .$$

Now, the loss function is defined as $\ell(k,y) = y_k$ for all $y \in \mathcal{Y}$ and $k \in \mathbb{N}$.

We construct a random outcome sequence and show that the expected value of the regret (with respect both to the random choice of the outcome sequence and to the forecaster's possibly random choices) for any possibly randomized forecaster is bounded from below by the claimed quantity.

More precisely, we denote by $U_1, \ldots, U_n$ the auxiliary randomization which the forecaster has access to. Without loss of generality, this sequence can be taken as an i.i.d. sequence of uniformly distributed random variables over $[0,1]$. Our underlying probability space is equipped with the σ-algebra of events generated by the random outcome sequence $Y_1, \ldots, Y_n$ and by the randomization $U_1, \ldots, U_n$. As the random outcome sequence is independent of the auxiliary randomization, we define N different probability distributions, $\mathbb{P}_i \otimes \mathbb{P}_A$, $i = 1, \ldots, N$, formed by the product of the auxiliary randomization (whose associated probability distribution is denoted by $\mathbb{P}_A$) and one of the N different probability distributions $\mathbb{P}_1, \ldots, \mathbb{P}_N$ over the outcome sequence defined as follows.

For $i = 1, \ldots, N$, $\mathbb{Q}_i$ is defined as the distribution (over $[0,1]$) of

$$Z^* 2^{-i} + \sum_{k=1,\ldots,N, \ k \neq i} Z_k 2^{-k} + 2^{-(N+1)} U \ ,$$

where U, Z^*, $Z_1, \ldots, Z_N$ are independent random variables such that U has uniform distribution, and Z^* and the Z_k have Bernoulli distribution with parameter $1/2 - \varepsilon$ for Z^* and $1/2$ for the Z_k. Now, the randomization is such that under $\mathbb{P}_i$, the outcome sequence $Y_1, \ldots, Y_n$ is i.i.d. with common distribution $\mathbb{Q}_i$.

Then, under each $\mathbb{P}_i$ (for $i = 1, \ldots, N$), the losses $\ell(k, Y_t)$, $k = 1, \ldots, N$, $t = 1, \ldots, n$, are independent Bernoulli random variables with the following parameters. For all t, $\ell(i, Y_t) = 1$ with probability $1/2 - \varepsilon$ and $\ell(k, Y_t) = 1$ with probability $1/2$ for each $k \neq i$, where ε is a positive number specified below.

We have

$$\begin{aligned}
&\max_{y_1,\ldots,y_n} \left(\mathbb{E}_A \widehat{L}_n - \min_{i=1,\ldots,N} L_{i,n} \right) \\
&= \max_{y_1,\ldots,y_n} \max_{i=1,\ldots,N} \left(\mathbb{E}_A \widehat{L}_n - L_{i,n} \right) \\
&\ge \max_{i=1,\ldots,N} \mathbb{E}_i \left[\mathbb{E}_A \widehat{L}_n - L_{i,n} \right] \ ,
\end{aligned}$$

where $\mathbb{E}_i$ (resp. $\mathbb{E}_A$) denotes expectation with respect to $\mathbb{P}_i$ (resp. $\mathbb{P}_A$).

Now, we use the following decomposition lemma, which states that a randomized algorithm performs, on the average, just as a convex combination of deterministic algorithms. The simple proof is omitted.

Lemma 14: For any randomized forecaster there exists an integer D, a point $\alpha = (\alpha_1, \ldots, \alpha_D) \in \mathbb{R}^D$ in the probability simplex, and D deterministic algorithms (indexed by a superscript $d = 1, \ldots, D$) such that, for every t and every possible outcome sequence $y_1^{t-1} = (y_1, \ldots, y_{t-1})$,

$$\mathbb{P}_A \left[I_t = i \,|\, y_1^{t-1} \right] = \sum_{d=1}^{D} \alpha_d \, \mathbb{I}_{[I_t^d = i \,|\, y_1^{t-1}]} \ ,$$

where $\mathbb{I}_{[I_t^d = i \,|\, y_1^{t-1}]}$ is the indicator function that the d-th deterministic algorithm chooses action i when the sequence of past outcomes is formed by y_1^{t-1}.

Using this lemma, we have that there exist D, α and D deterministic sub-algorithms such that

$$\begin{aligned}
&\max_{i=1,\ldots,N} \mathbb{E}_i \Big[\mathbb{E}_A \widehat{L}_n - L_{i,n} \Big] \\
&= \max_{i=1,\ldots,N} \mathbb{E}_i \left[\sum_{t=1}^{n} \sum_{d=1}^{D} \alpha_d \sum_{k=1}^{N} \mathbb{I}_{[I_t^d = k \,|\, Y_1^{t-1}]} \ell(k, Y_t) - L_{i,n} \right] \\
&= \max_{i=1,\ldots,N} \sum_{d=1}^{D} \alpha_d \, \mathbb{E}_i \left[\sum_{t=1}^{n} \sum_{k=1}^{N} \mathbb{I}_{[I_t^d = k \,|\, Y_1^{t-1}]} \ell(k, Y_t) - L_{i,n} \right]
\end{aligned}$$

Now, under $\mathbb{P}_i$ the regret grows by ε whenever an action different from i is chosen and remains the same otherwise. Hence,

$$\begin{aligned}
&\max_{i=1,\ldots,N} \mathbb{E}_i \Big[\mathbb{E}_A \widehat{L}_n - L_{i,n} \Big] \\
&= \max_{i=1,\ldots,N} \sum_{d=1}^{D} \alpha_d \, \mathbb{E}_i \left[\sum_{t=1}^{n} \sum_{k=1}^{N} \mathbb{I}_{[I_t^d = k \,|\, Y_1^{t-1}]} \ell(k, Y_t) - L_{i,n} \right] \\
&= \varepsilon \max_{i=1,\ldots,N} \sum_{d=1}^{D} \alpha_d \sum_{t=1}^{n} \mathbb{P}_i \left[I_t^d \neq i \right] \\
&= \varepsilon \, n \left(1 - \min_{i=1,\ldots,N} \sum_{d=1}^{D} \sum_{t=1}^{n} \frac{\alpha_d}{n} \mathbb{P}_i [I_t^d = i] \right) \ .
\end{aligned}$$

For the d-th deterministic subalgorithm, let $1 \le T_1^d < \ldots < T_m^d \le n$ be the times when the m queries were issued. Then $T_1^d, \ldots, T_m^d$ are finite stopping times with respect to the i.i.d. process $Y_1, \ldots, Y_n$. Hence, by a well-known fact in probability theory (see, e.g., [19, Lemma 2, page 138]), the revealed outcomes $Y_{T_1^d}, \ldots, Y_{T_m^d}$ are independent and indentically distributed as Y_1.

Let R_t^d be the number of revealed outcomes at time t and note that R_t^d is measurable with respect to the random outcome sequence. Now, as the subalgorithm we consider is deterministic, R_t^d is fully determined by $Y_{T_1^d}, \ldots, Y_{T_m^d}$. Hence, I_t^d may be seen as a function of $Y_{T_1^d}, \ldots, Y_{T_m^d}$ rather than a function of $Y_{T_1^d}, \ldots, Y_{T_{R_t^d}^d}$ only. As the joint distribution of $Y_{T_1^d}, \ldots, Y_{T_m^d}$ under $\mathbb{P}_i$ is $\mathbb{Q}_i^m$, we have proved that

$$\mathbb{P}_i [I_t^d = i] = \mathbb{Q}_i^m [I_t^d = i] \ .$$

Consequently, the lower bound rewrites as

$$\begin{aligned}&\max_{i=1,\ldots,N} \mathbb{E}_i \left[\mathbb{E}_A \widehat{L}_n - L_{i,n}\right] \\ &\quad = \varepsilon\, n \left(1 - \min_{i=1,\ldots,N} \sum_{d=1}^{D}\sum_{t=1}^{n} \frac{\alpha_d}{n} \mathbb{Q}_i^m [I_t^d = i]\right) .\end{aligned}$$

By the generalized Fano's inequality (see Lemma 18 in the Appendix), it is guaranteed that

$$\min_{i=1,\ldots,N} \sum_{d=1}^{D}\sum_{t=1}^{n} \frac{\alpha_d}{n} \mathbb{Q}_i^m [I_t^d = i] \le \max\left\{\frac{e}{1+e}, \frac{\bar{K}}{\ln(N-1)}\right\} ,$$

where

$$\begin{aligned}\bar{K} &= \sum_{t=1}^{n}\sum_{d=1}^{D}\sum_{i=2}^{N} \frac{\alpha_d}{n(N-1)} \mathrm{KL}(\mathbb{Q}_i^m, \mathbb{Q}_1^m) \\ &= \frac{1}{N-1}\sum_{i=2}^{N} \mathrm{KL}(\mathbb{Q}_i^m, \mathbb{Q}_1^m) ,\end{aligned}$$

and KL is the Kullback-Leibler divergence (or relative entropy) between two probability distributions.

Moreover, $\mathbb{B}_p$ denoting the Bernoulli distribution with parameter p,

$$\begin{aligned}&\mathrm{KL}(\mathbb{Q}_i^m, \mathbb{Q}_1^m) \\ &\quad = m\, \mathrm{KL}(\mathbb{Q}_i, \mathbb{Q}_1) \\ &\quad \le m \left(\mathrm{KL}\left(\mathbb{B}_{1/2-\varepsilon}, \mathbb{B}_{1/2}\right) + \mathrm{KL}\left(\mathbb{B}_{1/2}, \mathbb{B}_{1/2-\varepsilon}\right)\right) \\ &\quad = m\,\varepsilon \ln\left(1 + \frac{4\varepsilon}{1-2\varepsilon}\right) \le 5m\,\varepsilon^2\end{aligned}$$

for $0 \le \varepsilon \le 1/10$, where the first inequality holds by noting that the definition of the $\mathbb{Q}_i$ implies that the considered Kullback-Leibler divergence is upper bounded by the Kullback-Leibler divergence between $(Z_1, \ldots, Z^*, \ldots, Z_n, U)$, where Z^* is in the i-th position, and $(Z^*, Z_2 \ldots, Z_n, U)$. Therefore,

$$\begin{aligned}&\max_{y_1,\ldots,y_n} \left(\mathbb{E}_A \widehat{L}_n - \min_{i=1,\ldots,N} L_{i,n}\right) \\ &\quad \ge \varepsilon\, n \left(1 - \max\left\{\frac{e}{1+e}, \frac{5m\,\varepsilon^2}{\ln(N-1)}\right\}\right) .\end{aligned}$$

The choice

$$\varepsilon = \sqrt{\frac{e \ln(N-1)}{5(1+e)m}}$$

yields the claimed bound. ∎

APPENDIX I
BERNSTEIN'S INEQUALITY FOR MARTINGALES

We recall first a version of Bernstein's inequality suited for maxima of martingale difference sequences [20], and prove a corollary tailored to the needs of Section IV.

Lemma 15 (Bernstein's maximal inequality for martingales): Let $X_1, \ldots, X_n$ be a bounded martingale difference sequence with respect to the filtration $\mathcal{F} = (\mathcal{F}_t)_{1 \le t \le n}$ and with increments bounded in absolute values by K. Let

$$M_t = \sum_{s=1}^{t} X_s$$

be the associated martingale. Denote the sum of the conditional variances by

$$V_n = \sum_{t=1}^{n} \mathbb{E}\left[X_t^2 \,|\, \mathcal{F}_{t-1}\right] .$$

Then, for all $\lambda > 0$,

$$\left(\exp\left(\lambda M_n - \phi_K(\lambda) V_n\right)\right)_{n \ge 0}$$

is a supermartingale (with respect to the same filtration $\mathcal{F}$), where

$$\phi_K(\lambda) = \frac{1}{K^2}\left(e^{\lambda K} - 1 - \lambda K\right) .$$

In particular, for all constants $x, v > 0$,

$$\mathbb{P}\left[\max_{t=1,\ldots,n} M_t > x \text{ and } V_n \le v\right] \le \exp\left(-\frac{x^2}{2\,(v + Kx/3)}\right)$$

and therefore,

$$\mathbb{P}\left[\max_{t=1,\ldots,n} M_t > \sqrt{2vx} + (\sqrt{2}/3)Kx \;\text{ and }\; V_n \le v\right] \le e^{-x} .$$

Corollary 16: Under the assumptions of Lemma 15, for all $\delta \in (0,1)$, with probability at least $1-\delta$,

$$\max_{t=1,\ldots,n} M_t \le \sqrt{2(V_n + K^2)\ln(n/\delta)} + (\sqrt{2}/3)K\ln(n/\delta) .$$

PROOF. Denote

$$M = \max_{t=1,\ldots,n} M_t .$$

We apply the previous lemma n times and use a union-of-events bound. For $t = 1, \ldots, n$,

$$\begin{aligned}&\mathbb{P}\Big[M > \sqrt{2(V_n + K^2)\ln(n/\delta)} + (\sqrt{2}/3)K\ln(n/\delta) \\ &\qquad\qquad \text{and } \;V_n \in K^2\,[t-1,\,t]\Big] \\ &\quad \le \mathbb{P}\Big[M > \sqrt{2K^2 t\ln(n/\delta)} + (\sqrt{2}/3)K\ln(n/\delta) \\ &\qquad\qquad \text{and } \;V_n \le K^2 t\Big] \\ &\quad \le \delta/n ,\end{aligned}$$

where we used Lemma 15 in the last step. By boundedness of the X_t, V_n lies between 0 and $K^2\, n$, and therefore a union-of-events bound over $t = 1, \ldots, n$ concludes the proof. ∎

APPENDIX II
GENERALIZED FANO'S LEMMA

The crucial point in the proof of the lower bound theorem is an extension of Fano's lemma to a convex combination of probability masses, which may be proved thanks to a straightforward modification of the techniques developed by Birgé [21] (see also Massart [22]). Recall first a consequence of the variational formula for entropy.

Lemma 17: For arbitrary probability distributions $\mathbb{P}, \mathbb{Q}$ and for each $\lambda > 0$,

$$\lambda\, \mathbb{P}[A] - \psi_{\mathbb{Q}[A]}(\lambda) \le \mathrm{KL}(\mathbb{P}, \mathbb{Q})$$

where $\psi_p(\lambda) = \ln\big(p\,(e^\lambda - 1) + 1\big)$.

Lemma 18 (Generalized Fano): Let $\{A_{s,j} \,:\, s = 1,\ldots,S,\; j = 1,\ldots,N\}$ be a family of subsets of a set Ω such that $A_{s,1},\ldots,A_{s,N}$ form a partition of Ω for each fixed s. Let $\alpha_1,\ldots,\alpha_s$ be such that $\alpha_s \geq 0$ for $s = 1,\ldots,S$ and $\alpha_1 + \ldots + \alpha_S = 1$. Then, for all sets $\mathbb{P}_{s,1},\ldots,\mathbb{P}_{s,N}$, $s = 1,\ldots,S$, of probability distributions on Ω,

$$\min_{j=1,\ldots,N} \sum_{s=1}^{S} \alpha_s\, \mathbb{P}_{s,j}[A_{s,j}] \leq \max\left\{ \frac{e}{1+e}, \frac{\bar{K}}{\ln(N-1)} \right\} ,$$

where

$$\bar{K} = \sum_{s=1}^{S}\sum_{j=2}^{N} \frac{\alpha_s}{N-1}\, \mathrm{KL}(\mathbb{P}_{s,j}, \mathbb{P}_{s,1}) \;.$$

PROOF. Using Lemma 17, we have that

$$\sum_{s=1}^{S}\sum_{j=2}^{N} \frac{\alpha_s}{N-1}\lambda\, \mathbb{P}_{s,j}[A_{s,j}] - \sum_{s=1}^{S}\sum_{j=2}^{N} \frac{\alpha_s}{N-1}\psi_{\mathbb{P}_{s,1}[A_{s,j}]}(\lambda) \leq \sum_{s=1}^{S}\sum_{j=2}^{N} \frac{\alpha_s}{N-1}\mathrm{KL}(\mathbb{P}_{s,j}, \mathbb{P}_{s,1}) = \bar{K} \;.$$

Now, for each fixed $\lambda > 0$, the function that maps p to $-\psi_p(\lambda)$ is convex. Hence, letting

$$\begin{aligned} p_1 &= \sum_{s=1}^{S}\sum_{j=2}^{N} \frac{\alpha_s}{N-1}\mathbb{P}_{s,1}[A_{s,j}] \\ &= \frac{1}{N-1}\left(1 - \sum_{s=1}^{S} \alpha_s \mathbb{P}_{s,1}[A_{s,1}]\right) , \end{aligned}$$

by Jensen's inequality we get

$$\begin{aligned} &\sum_{s=1}^{S}\sum_{j=2}^{N} \frac{\alpha_s}{N-1}\lambda\, \mathbb{P}_{s,j}[A_{s,j}] - \psi_{p_1}(\lambda) \\ \leq\;& \sum_{s=1}^{S}\sum_{j=2}^{N} \frac{\alpha_s}{N-1}\lambda\, \mathbb{P}_{s,j}[A_{s,j}] - \sum_{s=1}^{S}\sum_{j=2}^{N} \frac{\alpha_s}{N-1}\psi_{\mathbb{P}_{s,1}[A_{s,j}]}(\lambda) \;. \end{aligned}$$

Recalling that the right-hand side of the above inequality above is less than $\bar{K}$, and introducing the quantities

$$a_j = \sum_{s=1}^{S} \alpha_s \mathbb{P}_{s,j}[A_{s,j}] \qquad \text{for } j = 1,\ldots,N,$$

we conclude

$$\lambda \min_{j=1,\ldots,N} a_j - \psi_{\frac{1-a_1}{N-1}}(\lambda) \leq \lambda \frac{1}{N-1}\sum_{j=2}^{N} a_j - \psi_{\frac{1-a_1}{N-1}}(\lambda) \leq \bar{K} \;.$$

Denote by a the minimum of the a_j's and let $p^* = (1-a)/(N-1) \geq p_1$. We only have to deal with the case when $a \geq e/(1+e)$. As for all $\lambda > 0$, the function that maps p to $-\psi_p$ is decreasing, we have

$$\begin{aligned} \bar{K} &\geq \sup_{\lambda > 0}\big(\lambda\, a - \psi_{p^*}(\lambda)\big) \geq a \ln \frac{a}{e\, p^*} \\ &\geq a \ln \frac{a\,(N-1)}{(1-a)\,e} \geq a \ln(N-1) \;, \end{aligned}$$

whenever $p^* \leq a \leq 1$ for the second inequality to hold, and by using $a \geq e/(1+e)$ for the last one. As $p^* \leq 1/(N-1) \leq e/(1+e)$ whenever $N \geq 3$, the case $a < p^*$ may only happen when $N = 2$, but then the result is trivial. ∎

APPENDIX III
A BASIC FACT

Lemma 19: If $x_t, y_t \geq 0$, and $b \geq 0$, are such that for all $t = 1,\ldots,n$

$$x_t \leq y_t + b\sqrt{x_n} \;, \tag{10}$$

then

$$\forall t = 1,\ldots,n \qquad x_t \leq y_t + b\sqrt{y_n} + b^2 \;.$$

PROOF. We obtain a bound over $\sqrt{x_n}$ and apply it to (10) to conclude. The inequality

$$x_n \leq y_n + b\sqrt{x_n}$$

rewrites as

$$\left(\sqrt{x_n} - \frac{b}{2}\right)^2 \leq y_n + \frac{b^2}{4} \;,$$

that is, either $\sqrt{x_n} \leq b/2$ or

$$\sqrt{x_n} - \frac{b}{2} = \left|\sqrt{x_n} - \frac{b}{2}\right| \leq \sqrt{y_n + \frac{b^2}{4}} \leq \sqrt{y_n} + \frac{b}{2} \;.$$

In both cases,

$$\sqrt{x_n} \leq b + \sqrt{y_n}$$

concluding the proof. ∎

REFERENCES

[1] J. Hannan, "Approximation to Bayes risk in repeated play," *Contributions to the theory of games*, vol. 3, pp. 97–139, 1957.

[2] D. Blackwell, "Controlled random walks," in *Proceedings of the International Congress of Mathematicians, 1954*, vol. III. Amsterdam: North-Holland, 1956, pp. 336–338.

[3] N. Cesa-Bianchi, Y. Freund, D. Helmbold, D. Haussler, R. Schapire, and M. Warmuth, "How to use expert advice," *Journal of the ACM*, vol. 44, no. 3, pp. 427–485, 1997.

[4] D. Helmbold and S. Panizza, "Some label efficient learning results," in *Proceedings of the 10th Annual Conference on Computational Learning Theory*. ACM Press, 1997, pp. 218–230.

[5] J.-F. Mertens, S. Sorin, and S. Zamir, "Repeated games," CORE Discussion paper, no. 9420,9421,9422, Louvain-la-Neuve, 1994.

[6] A. Rustichini, "Minimizing regret: The general case," *Games and Economic Behavior*, vol. 29, pp. 224–243, 1999.

[7] A. Piccolboni and C. Schindelhauer, "Discrete prediction games with arbitrary feedback and loss," in *Proceedings of the 14th Annual Conference on Computational Learning Theory*, 2001, pp. 208–223.

[8] S. Mannor and N. Shimkin, "On-line learning with imperfect monitoring," in *Proceedings of the 16th Annual Conference on Learning Theory*. Springer, New York, 2003, pp. 552–567.

[9] N. Cesa-Bianchi, G. Lugosi, and G. Stoltz, "Regret minimization under partial monitoring," 2004, manuscript.

[10] A. Baños, "On pseudo-games," *Annals of Mathematical Statistics*, vol. 39, pp. 1932–1945, 1968.

[11] N. Megiddo, "On repeated games with incomplete information played by non-Bayesian players," *International Journal of Game Theory*, vol. 9, pp. 157–167, 1980.

[12] D. Foster and R. Vohra, "Asymptotic calibration," *Biometrika*, vol. 85, pp. 379–390, 1998.

[13] S. Hart and A. Mas-Colell, "A reinforcement procedure leading to correlated equilibrium," in *Economic Essays: A Festschrift for Werner Hildenbrand*, G. Debreu, W. Neuefeind, and W. Trockel, Eds. New York: Springer, 2002, pp. 181–200.

[14] P. Auer, N. Cesa-Bianchi, Y. Freund, and R. Schapire, "The nonstochastic multiarmed bandit problem," *SIAM Journal on Computing*, vol. 32, pp. 48–77, 2002.

[15] P. Auer, "Using confidence bounds for exploitation-exploration tradeoffs," *Journal of Machine Learning Research*, vol. 3, pp. 397–422, 2002, a preliminary version has appeared in *Proc. of the 41th Annual Symposium on Foundations of Computer Science*.

[16] D. Helmbold, N. Littlestone, and P. Long, "Apple tasting," *Information and Computation*, vol. 161, no. 2, pp. 85–139, 2000.

[17] P. Auer, N. Cesa-Bianchi, Y. Freund, and R. Schapire, "The nonstochastic multiarmed bandit problem," *SIAM Journal on Computing*, vol. 32, no. 1, 2002.

[18] W. Hoeffding, "Probability inequalities for sums of bounded random variables," *Journal of the American Statistical Association*, vol. 58, pp. 13–30, 1963.

[19] Y. Chow and H. Teicher, *Probability Theory*. Springer, 1988.

[20] D. Freedman, "On tail probabilities for martingales," *Annals of Probability*, vol. 3, pp. 100–118, 1975.

[21] L. Birgé, "A new look at an old result: Fano's lemma," Technical report, Université Paris 6. 2001.

[22] P. Massart. *"Concentration inequalities and model selection,"* Saint-Flour summer school lecture notes, 2003. To appear.

PLACE
PHOTO
HERE

Nicolò Cesa-Bianchi was born on June 18, 1963 in Milan, Italy. He received his Ph.D. in computer science from the University of Milan in 1993. Since 2003 he has been full professor of computer science at the University of Milan. His research interests are in the areas of machine learning and statistical pattern recognition.

Since 2002 he is associate editor for the Journal of Machine Learning, Kluwer Academic Publishers.

PLACE
PHOTO
HERE

Gábor Lugosi was born on July 13, 1964 in Budapest, Hungary. He graduated in electrical engineering at the Technical University of Budapest in 1987, and received his Ph.D. from the Hungarian Academy of Sciences in 1991. Since September 1996, he has been at the Department of Economics, Pompeu Fabra University. His research interest involves pattern recognition, nonparametric statistics and information theory.

He is a co-author of the monographs L.Devroye, L. Györfi, G.Lugosi, "A probabilistic theory of pattern recognition," Springer, New York, 1996 and L.Devroye, G.Lugosi, "Combinatorial methods of density estimation" Springer, New York, 2000.

PLACE
PHOTO
HERE

Gilles Stoltz was born on May 27, 1979 in Wissembourg, France. He graduated in mathematics at the Ecole Normale Supérieure, Cachan, in 2002, and is to receive his Ph.d. from University Paris-Sud, Orsay, in 2005. Since September 2004, he has been at the Departement of Mathematics and Applications, Ecole Normale Supérieure, Paris. His research interest involves machine learning from a statistical viewpoint.