COMS 4773: Online allocation

Daniel Hsu

February 17, 2024

1 Online allocation

The Online Allocation problem, proposed by Freund and Schapire (1997), is a generalization of the "Using Expert Advice" problem under a particular restriction of Nature that will be explained shortly. Define the following sets of N-vectors:

$$\mathbb{R}^N_+ := \big\{x = (x_1, \dots, x_N) \in \mathbb{R}^N : x_i \geq 0 \text{ for all } i \in [N]\big\},$$
$$\Delta^{N-1} := \left\{x = (x_1, \dots, x_N) \in \mathbb{R}^N_+ : \sum_{i=1}^N x_i = 1\right\}.$$

(The set $\mathbb{R}^N_+$ is the non-negative orthant of $\mathbb{R}^N$, and the set Δ^{N-1} is the probability simplex in $\mathbb{R}^N$.) And define the standard inner product $\langle \cdot, \cdot \rangle : \mathbb{R}^N \times \mathbb{R}^N \to \mathbb{R}$ by

$$\langle x, y \rangle := \sum_{i=1}^N x_i y_i \quad \text{for all } x = (x_1, \dots, x_N) \in \mathbb{R}^N \text{ and } y = (y_1, \dots, y_N) \in \mathbb{R}^N.$$

For round $t = 1, 2, \dots$:

- The learner chooses an allocation vector $p_t \in \Delta^{N-1}$.
- Nature then reveals a loss vector $\ell_t \in \mathbb{R}^N_+$ to the learner.
- The learner incurs loss $\langle p_t, \ell_t \rangle$.

The cumulative (or total) loss of the learner after T rounds is

$$L_T := \sum_{t=1}^T \langle p_t, \ell_t \rangle.$$

(The number T is called the *horizon*.) The cumulative loss of the allocation vector $q \in \Delta^{N-1}$ is

$$L_{T,q} := \sum_{t=1}^T \langle q, \ell_t \rangle.$$

If q is an elementary vector[1], then we write $L_{T,i}$ to mean L_{T,e_i}. The *regret of the learner to q after T rounds* is

$$R_{T,q} := L_T - L_{T,q}$$

[1]The N *elementary (or coordinate) vectors* (in $\mathbb{R}^N$) are $e_1, \dots, e_N$, where e_i has a 1 in the i-th entry and 0 elsewhere.

(with $R_{T,i} = R_{T,e_i}$), and

$$R_T := \max_{q \in \Delta^{N-1}} R_{T,q} = L_T - \min_{q \in \Delta^{N-1}} L_{T,q}$$

is the *regret of the learner (to the best allocation) after T rounds*.

Regret is named as such because it measures the extra loss incurred by the learner compared to having used the "best-in-hindsight" allocation in all T rounds. Note that, by linearity, we have

$$\sum_{t=1}^{T} \langle q, \ell_t \rangle = \left\langle q, \sum_{t=1}^{T} \ell_t \right\rangle,$$

so the "best-in-hindsight" cumulative loss achievable by a single allocation vector is determined by minimizing a linear function (given by $\sum_{t=1}^{T} \ell_t$) over the probability simplex. This maximum value will always be achieved at a vertex of the simplex, i.e., one of the N elementary vectors. So R_T can also be written as

$$R_T = \max_{i \in [N]} L_T - L_{T,i} = L_T - \min_{i \in [N]} L_{T,i}.$$

2 Connection between online allocation and using expert advice

Consider the following randomized version of WEIGHTED MAJORITY for the "Using Expert Advice" problem, also due to Littlestone and Warmuth (1994), and aptly named RANDOMIZED WEIGHTED MAJORITY. The weights $w_{t,i}$ are initialized and updated in exactly the same way as in WEIGHTED MAJORITY; the only difference is in how the prediction is determined.

- Let $p_t = (p_{t,1}, \ldots, p_{t,N}) \in \Delta^{N-1}$ be defined by

  $$p_{t,i} := \frac{w_{t,i}}{Z_t} \quad \text{for all } i \in [N],$$

  where $Z_t := \sum_{i=1}^{N} w_{t,i}$.

- Let i_t be a $[N]$-valued random variable that that takes value $i \in [N]$ with probability $p_{t,i}$.

- The learner predicts $a_t = b_{t,i_t}$.

Note that this algorithm is still subject to the lower bound of Cover (1965).

However, if we further restrict Nature in the following way, then we will be able to draw a connection between the way RANDOMIZED WEIGHTED MAJORITY makes predictions for the "Using Expert Advice" problem and the Online Allocation problem. Specifically, we assume that Nature is an *oblivious adversary*[2], which means the following: Nature fixes, once and for all prior to the first round, the (infinite) sequence of outcomes $y_1, y_2, \ldots$, as well as the (infinite) sequences of the experts' predictions $b_{1,i}, b_{2,i}, \ldots$ for each $i \in [N]$. This assumption prevents Cover's scenario, since Nature now cannot simply set y_t to be the opposite of the learner's prediction.

[2]RANDOMIZED WEIGHTED MAJORITY also works well under slightly weaker assumptions on Nature. Specifically, it suffices to allow Nature to choose the experts' predictions $b_{t,i}$'s and the outcome y_t after the learner choose p_t, but these choices must be made before the random choice of i_t is realized.

Observe that p_t in RANDOMIZED WEIGHTED MAJORITY is a deterministic function of the outcomes and expert's predictions. Therefore, the probability that the learner makes a mistake in round t is

$$\begin{aligned}
\Pr(a_t \neq y_t) &= \sum_{i=1}^{N} \Pr(a_t \neq y_t \wedge i_t = i) \\
&= \sum_{i=1}^{N} \Pr(i_t = i) \cdot \Pr(a_t \neq y_t \mid i_t = i) \\
&= \sum_{i=1}^{N} \Pr(i_t = i) \cdot \Pr(b_{t,i} \neq y_t \mid i_t = i) \\
&= \sum_{i=1}^{N} p_{t,i} \cdot \mathbb{1}\{b_{t,i} \neq y_t\} \\
&= \langle p_t, \ell_t \rangle
\end{aligned}$$

where we define the loss vector $\ell_t \in \mathbb{R}_+^N$ by

$$\ell_t := (\mathbb{1}\{b_{t,1} \neq y_t\}, \ldots, \mathbb{1}\{b_{t,N} \neq y_t\}) \in [0,1]^N. \tag{1}$$

Therefore, we can express the sum of these probabilities in terms of quantities from the Online Allocation problem:

$$\sum_{t=1}^{T} \Pr(a_t \neq y_t) = \sum_{t=1}^{T} \langle p_t, \ell_t \rangle = L_T.$$

This sum of probabilities is also equal to the expected number of mistakes made by the learner after T rounds.

With this definition of loss vectors ℓ_t, it is also easy to verify that $L_{T,i}$ is the number of mistakes of the i-th expert after T rounds. And therefore the regret $R_T = L_T - \min_{i \in [N]} L_{T,i}$ is the expected number of additional mistakes of the learner beyond the number of mistakes of the best expert.

3 Hedge

HEDGE is an algorithm due to (Freund and Schapire, 1997) for the Online Allocation problem.

- Initially, set $w_{1,i} := 1$ for all $i \in [N]$.
- In round t:
  - Let $p_t = (p_{t,1}, \ldots, p_{t,N}) \in \Delta^{N-1}$ be defined by

    $$p_{t,i} := \frac{w_{t,i}}{Z_t} \quad \text{for all } i \in [N],$$

    where $Z_t := \sum_{i=1}^{N} w_{t,i}$.
  - Observe loss vector $\ell_t = (\ell_{t,1}, \ldots, \ell_{t,N}) \in \mathbb{R}_+^N$.

– Update: for each $i \in [N]$,

$$w_{t+1,i} = w_{t,i} \cdot \exp(-\eta \ell_{t,i}).$$

Above, $\eta > 0$ is a hyperparameter of the algorithm. (Freund and Schapire (1997) use the parameterization $\beta = e^{-\eta}$.) It is clear that HEDGE updates the weights $w_{t,i}$ in exactly the same way as (RANDOMIZED) WEIGHTED MAJORITY for the special case of loss vectors from (1).

Theorem 1 (Freund and Schapire, 1997)**.** *For any T, any $\eta > 0$, and any loss vectors $\ell_1, \dots, \ell_T \in \mathbb{R}_+^N$, HEDGE with hyperparameter $\eta > 0$ guarantees that*

$$R_T \le \frac{\log N}{\eta} + \frac{\eta}{2} \sum_{t=1}^{T} \langle p_t, \ell_t^2 \rangle$$

where $\ell_t^2 := (\ell_{t,1}^2, \dots, \ell_{t,N}^2) \in \mathbb{R}_+^N$ for all t. Furthermore, if $\ell_t \in [0,1]^N$ for all t, then

$$R_T \le \frac{\log N}{\eta} + \frac{\eta T}{2} \quad \textit{and} \quad (1 - \eta/2) R_T \le \frac{\log N}{\eta} + \frac{\eta}{2} \min_{i \in [N]} L_{T,i}.$$

Finally, there is a choice of $\eta > 0$ such that $R_T \le \sqrt{2T \log N}$, and there is also a choice of $\eta > 0$ such that $R_T \le O(\sqrt{\min_{i \in [N]} L_{T,i} \log N} + \log N)$.

The statement of Theorem 1 is slightly different from how it is stated by Freund and Schapire (1997) but not in any essential way for our purposes. The proof below is based on that from Freund and Schapire (1999), which uses the concept of relative entropy (a.k.a. Kullback-Leibler divergence).

4 Entropy and relative entropy

The *(Shannon) entropy* $\mathrm{H}(q)$ of a (discrete) probability distribution $q = (q_1, \dots, q_N) \in \Delta^{N-1}$ over $[N]$ is defined by

$$\mathrm{H}(q) := \sum_{i=1}^{N} q_i \log \frac{1}{q_i}.$$

(Here, we regard $0 \log 0$ as 0 rather than $-\infty$.) Entropy is a central concept in information theory and many other fields, with several interpretations. One way to think about entropy is as a measure of average "surprise" one experiences when observing the value of a random variable. The idea is that you experience more surprise when you see rare events than when you see common events. Letting $\log(1/q_i)$ be the quantitative measure of surprise for the outcome i, we then see $\mathrm{H}(q)$ to be exactly the expected surprise. (There are many other interpretations of entropy, such as the average number of bits needed to communicate a random message $X \sim q$, but we will not discuss them here.) The maximum value of $H(q)$ is $\log N$, achieved by the uniform distribution $(1/N, \dots, 1/N)$.

Now consider two probability distributions, $p, q \in \Delta^{N-1}$. The *cross entropy* of q from p is

$$\mathrm{CE}(q, p) := \sum_{i=1}^{N} q_i \log \frac{1}{p_i}.$$

Cross entropy measures the average surprise when outcomes are distributed according to q, but the measure of surprise is (perhaps mistakenly) based on p. The *relative entropy* of q from p is

$$\mathrm{RE}(q,p) := \sum_{i=1}^{N} q_i \log \frac{q_i}{p_i} = \mathrm{CE}(q,p) - \mathrm{H}(q)$$

is the excess average surprise in this scenario. Note that $\mathrm{RE}(q,p)$ is well-defined only if q is dominated by p (written $q \ll p$, which means that for any i, we have $p_i = 0$ only if $q_i = 0$ as well). If $q \not\ll p$, then we define $\mathrm{RE}(q,p) := +\infty$.

Proposition 1 (Gibbs' inequality)**.** $\mathrm{RE}(q,p) \geq 0$ *with equality if and only if* $q = p$.

Proof. A Taylor expansion of log gives

$$\log(1+x) \leq x \quad \text{for all } x > -1. \tag{2}$$

Without loss of generality, assume $q \ll p$. Then

$$\begin{aligned}
\mathrm{RE}(q,p) &= -\sum_{i=1}^{N} q_i \log\left(1 + \frac{p_i}{q_i} - 1\right) \\
&\geq -\sum_{i=1}^{N} q_i \left(\frac{p_i}{q_i} - 1\right) && \text{(by (2))} \\
&= -\sum_{i=1}^{N} p_i - q_i = 0.
\end{aligned}$$

Equality holds iff $p_i/q_i - 1 = 0$ for all i, which is equivalent to $p_i = q_i$ for all i. □

5 Proof of Theorem 1

To prove Theorem 1, it suffices to show that, for any $q \in \Delta^{N-1}$, we have

$$\sum_{t=1}^{T} \langle p_t - q, \ell_t \rangle \leq \frac{\log N}{\eta} + \frac{\eta}{2} \sum_{t=1}^{T} \langle p_t, \ell_t^2 \rangle.$$

The additional parts (after "Furthermore") follow by using the fact that $\ell_{t,i}^2 \leq \ell_{t,i}$ when $\ell_t \in [0,1]^N$.

The following is the key lemma in the proof.

Lemma 1. *Fix any* $p \in \Delta^{N-1}$ *and any* $\ell \in \mathbb{R}_+^N$*, and define* $p' \in \Delta^{N-1}$ *by*

$$p'_i := \frac{p_i \exp(-\eta \ell_i)}{\sum_{j=1}^{N} p_j \exp(-\eta \ell_j)} \quad \textit{for all } i \in [N].$$

Then for any $q \in \Delta^{N-1}$ *such that* $q \ll p$*,*

$$\mathrm{RE}(q,p') - \mathrm{RE}(q,p) \leq \eta \langle q - p, \ell \rangle + \frac{\eta^2}{2} \langle p, \ell^2 \rangle.$$

where $\ell^2 := (\ell_1^2, \ldots, \ell_N^2)$.

Proof. Taylor expansion of exp gives

$$\exp(x) \le 1 + x + \frac{x^2}{2} \quad \text{for all } x \le 0. \tag{3}$$

Therefore

$$\begin{aligned}
\mathrm{RE}(q, p') - \mathrm{RE}(q, p) = \sum_{i=1}^{N} q_i \log \frac{p_i}{p'_i} &= \eta\langle q, \ell\rangle + \log\left(\sum_{j=1}^{N} p_j \exp(-\eta\ell_j)\right) \\
&\le \eta\langle q, \ell\rangle + \log\left(\sum_{j=1}^{N} p_j\left(1 - \eta\ell_j + \frac{\eta^2}{2}\ell_j^2\right)\right) \qquad \text{(by (3))} \\
&= \eta\langle q, \ell\rangle + \log\left(1 - \eta\langle p, \ell\rangle + \frac{\eta^2}{2}\langle p, \ell^2\rangle\right) \\
&\le \eta\langle q, \ell\rangle - \eta\langle p, \ell\rangle + \frac{\eta^2}{2}\langle p, \ell^2\rangle \qquad \text{(by (2)).} \quad \square
\end{aligned}$$

Using Lemma 1, we have for any t and any $q \in \Delta^{N-1}$,

$$\eta\langle p_t - q, \ell_t\rangle \le \mathrm{RE}(q, p_t) - \mathrm{RE}(q, p_{t+1}) + \frac{\eta^2}{2}\langle p_t, \ell_t^2\rangle.$$

Sum both sides over $t = 1, \ldots, T$ to obtain

$$\begin{aligned}
\eta\sum_{t=1}^{T}\langle p_t - q, \ell_t\rangle &\le \sum_{t=1}^{T} \mathrm{RE}(q, p_t) - \mathrm{RE}(q, p_{t+1}) + \frac{\eta^2}{2}\langle p_t, \ell_t^2\rangle \\
&= \underbrace{\mathrm{RE}(q, p_1)}_{\le \log N} - \underbrace{\mathrm{RE}(q, p_{T+1})}_{\ge 0} + \frac{\eta^2}{2}\sum_{t=1}^{T}\langle p_t, \ell_t^2\rangle \\
&\le \log N + \frac{\eta^2}{2}\sum_{t=1}^{T}\langle p_t, \ell_t^2\rangle
\end{aligned}$$

where the last inequality uses Proposition 1 and the definition of p_1. $\square$

6 Application to zero-sum games

A zero-sum game is a two-player game in which the loss of one player is the gain of the other player. The loss/gain is specified by a payoff matrix $A \in \mathbb{R}^{m\times n}$; for simplicity, assume the entries of A are from $[0, 1]$. The "row" player has m possible actions, and the "column" player has n possible actions. The players are permitted to use *mixed strategies*, which are probability distributions over their respective action sets. If the row player chooses $p \in \Delta^{m-1}$ and the column player chooses $q \in \Delta^{n-1}$, then the loss of the row player (equivalently, the gain of the column player) is $p^{\mathsf{T}}Aq$.

We have been ambiguous about exactly when the choices of the players are made. If the row player chooses p first, and then the column player chooses q in response to seeing p, then the best possible loss of the row player is

$$\min_{p \in \Delta^{m-1}} \max_{q \in \Delta^{n-1}} p^\mathsf{T} A q. \tag{4}$$

However, if the column player chooses q first, and then the row player chooses p in response to seeing q, then the best possible gain of the column player is

$$\max_{q \in \Delta^{n-1}} \min_{p \in \Delta^{m-1}} p^\mathsf{T} A q. \tag{5}$$

Von Neumann's min-max theorem states that (4) and (5) are equal; there is no advantage to "going first" in a zero-sum game. This common value of (4) and (5) is called the *value of the game*.

Freund and Schapire (1999) showed that Von Neumann's min-max theorem can be proved largely as a corollary of Theorem 1. There are two parts to the proof. The first part is the "easy" direction (and doesn't involve Theorem 1), although we give the steps below very explicitly for completeness.

Proposition 2. *For any $A \in [0,1]^{m \times n}$,*

$$\max_{q \in \Delta^{n-1}} \min_{p \in \Delta^{m-1}} p^\mathsf{T} A q \leq \min_{p \in \Delta^{m-1}} \max_{q \in \Delta^{n-1}} p^\mathsf{T} A q.$$

Proof. Fix any $q' \in \Delta^{n-1}$. Then, for every $p \in \Delta^{m-1}$,

$$p^\mathsf{T} A q' \leq \max_{q \in \Delta^{n-1}} p^\mathsf{T} A q.$$

Let p' be the choice of $p \in \Delta^{m-1}$ that minimizes the right-hand side, so

$$\begin{aligned}(p')^\mathsf{T} A q' &\leq \max_{q \in \Delta^{n-1}} (p')^\mathsf{T} A q \\ &= \min_{p \in \Delta^{m-1}} \max_{q \in \Delta^{n-1}} p^\mathsf{T} A q.\end{aligned}$$

We clearly also have

$$(p')^\mathsf{T} A q' \geq \min_{p \in \Delta^{m-1}} p^\mathsf{T} A q',$$

so combining the previous two displayed inequalities gives

$$\min_{p \in \Delta^{m-1}} p^\mathsf{T} A q' \leq \min_{p \in \Delta^{m-1}} \max_{q \in \Delta^{n-1}} p^\mathsf{T} A q$$

Since this inequality holds for every $q' \in \Delta^{n-1}$, it also holds for the choice of q' that maximizes the left-hand side:

$$\max_{q \in \Delta^{n-1}} \min_{p \in \Delta^{m-1}} p^\mathsf{T} A q \leq \min_{p \in \Delta^{m-1}} \max_{q \in \Delta^{n-1}} p^\mathsf{T} A q. \qquad \square$$

The second part is the "hard" direction, which involves the execution of HEDGE in an instance of the Online Allocation problem that is constructed on-the-fly.

Proposition 3. *For any $A \in [0,1]^{m \times n}$,*

$$\min_{p \in \Delta^{m-1}} \max_{q \in \Delta^{n-1}} p^\mathsf{T} A q \leq \max_{q \in \Delta^{n-1}} \min_{p \in \Delta^{m-1}} p^\mathsf{T} A q.$$

Proof. Consider the execution of HEDGE for selecting $p_1, p_2, \ldots, p_T \in \Delta^{m-1}$ (so $N = m$) with loss vectors $\ell_1, \ell_2, \ldots, \ell_T \in [0,1]^m$ defined by $\ell_t = Aq_t$ for all t, where $q_t \in \Delta^{n-1}$ is chosen so that

$$p_t^\mathsf{T} A q_t = \max_{q \in \Delta^{n-1}} p_t^\mathsf{T} A q.$$

By Theorem 1, using $\eta = \sqrt{(2 \log m)/T}$, we have

$$\sum_{t=1}^{T} \langle p_t, \ell_t \rangle \leq \min_{p \in \Delta^{m-1}} \sum_{t=1}^{T} \langle p, \ell_t \rangle + \sqrt{2T \log m}.$$

This implies

$$\frac{1}{T} \sum_{t=1}^{T} p_t^\mathsf{T} A q_t \leq \min_{p \in \Delta^{m-1}} \frac{1}{T} \sum_{t=1}^{T} p^\mathsf{T} A q_t + \sqrt{\frac{2 \log m}{T}}. \tag{6}$$

Define $\bar{p} = \frac{1}{T} \sum_{t=1}^{T} p_t$ and $\bar{q} = \frac{1}{T} \sum_{t=1}^{T} q_t$. Then

$$\begin{aligned}
\min_{p \in \Delta^{m-1}} \max_{q \in \Delta^{n-1}} p^\mathsf{T} A q &\leq \max_{q \in \Delta^{n-1}} \bar{p}^\mathsf{T} A q \\
&= \max_{q \in \Delta^{n-1}} \frac{1}{T} \sum_{t=1}^{T} p_t^\mathsf{T} A q \\
&\leq \frac{1}{T} \sum_{t=1}^{T} \max_{q \in \Delta^{n-1}} p_t^\mathsf{T} A q \\
&= \frac{1}{T} \sum_{t=1}^{T} p_t^\mathsf{T} A q_t \\
&\leq \min_{p \in \Delta^{m-1}} \frac{1}{T} \sum_{t=1}^{T} p^\mathsf{T} A q_t + \sqrt{\frac{2 \log m}{T}} \quad \text{(by (6))} \\
&= \min_{p \in \Delta^{m-1}} p^\mathsf{T} A \bar{q} + \sqrt{\frac{2 \log m}{T}} \\
&\leq \max_{q \in \Delta^{n-1}} \min_{p \in \Delta^{m-1}} p^\mathsf{T} A q + \sqrt{\frac{2 \log m}{T}}.
\end{aligned}$$

Now let $T \to \infty$, to conclude the proof. □

The proof of Proposition 3 shows how to use HEDGE to construct the mixed strategies $\bar{p}$ and $\bar{q}$, essentially by repeatedly simulating the two-player game, where the row player's $p_t \in \Delta^{m-1}$ is controlled by HEDGE, and the column player's $q_t \in \Delta^{n-1}$ is the *best response* to p_t. If $\mathrm{val}(A)$ denotes the value of the game for payoff matrix A, then $\bar{p}$ and $\bar{q}$ satisfy

$$\max_{q \in \Delta^{n-1}} \bar{p}^\mathsf{T} A q \leq \mathrm{val}(A) + \sqrt{\frac{2 \log m}{T}} \quad \text{and} \quad \min_{p \in \Delta^{m-1}} p^\mathsf{T} A \bar{q} \geq \mathrm{val}(A) - \sqrt{\frac{2 \log m}{T}}.$$

The proof does not specifically rely on HEDGE (except for the form of the regret bound in (6)); other algorithms that ensure sublinear regret for Online Allocation could also be used.

7 Lower bound

How good is HEDGE for the Online Allocation problem? Its worst-case regret bound after T rounds is $O(\sqrt{T \log N})$ assuming loss vectors come from $[0,1]^N$. It turns out this is the best possible, for any algorithm (up to constants).

Theorem 2. *Consider any algorithm for the Online Allocation problem. Let $R_T(\ell_1, \ldots, \ell_T)$ denote the regret of the algorithm after T rounds with loss vectors $\ell_1, \ldots, \ell_T \in [0,1]^N$. Then*

$$\sup_{T,N \in \mathbb{N}} \sup_{\ell_1,\ldots,\ell_T \in [0,1]^N} \frac{R_T(\ell_1, \ldots, \ell_T)}{\sqrt{\frac{T \log N}{2}}} \geq 1.$$

To prove a lower bound of this sort, it would seem one needs to reason about how all possible algorithms for Online Allocation. This should seem like a rather daunting task. We will leverage the constraints imposed by the Online Allocation problem on what "information" the algorithm has access to when choosing its allocation vector in each round. How can this be enough to prove a lower bound?

We completely side-step the issue of reasoning about what loss vectors would make a particular algorithm incur large regret. This is achieved by using the *probabilistic method*. The idea is the following:

1. Define a probability distribution for the loss vectors $\ell_1, \ldots, \ell_T$, and compute (a lower bound on) the expected value of $R_T(\ell_1, \ldots, \ell_T)$.
2. Leverage the fact that there must be some element in the sample space, which ultimately determines $\ell_1, \ldots, \ell_T$ and $R_T(\ell_1, \ldots, \ell_T)$, such that $R_T(\ell_1, \ldots, \ell_T)$ is at least as large as its expected value.

So it suffices to show that

$$\mathbb{E}[R_T(\ell_1, \ldots, \ell_T)] \geq \sqrt{\frac{T \log N}{2}}$$

for some suitable probability distribution for $\ell_1, \ldots, \ell_T$.

(TO BE COMPLETED)

8 Multi-armed bandit problem

References

Thomas M Cover. Behavior of sequential predictors of binary sequences. *Transactions on Prague Conference on Information Theory Statistical Decision Functions, Random Processes*, pages 263–272, 1965.

Yoav Freund and Robert E Schapire. A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of Computer and System Sciences*, 55(1):119–139, 1997.

Yoav Freund and Robert E Schapire. Adaptive game playing using multiplicative weights. *Games and Economic Behavior*, 29(1-2):79–103, 1999.

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