Algorithms and Uncertainty, Summer 2020 — Lecture 19 (5 pages)

No-Regret Learning: Lower Bound

Thomas Kesselheim — Last Update: July 2, 2020

There are T steps and n actions. An adversary determines a sequence of cost vectors $\ell^{(1)}, \ldots, \ell^{(T)}$ in advance, defining the cost $\ell_i^{(t)} \in [0,1]$ of action i in step t. This sequence is initially unknown to the algorithm. In step t, the algorithm chooses one of the n actions, denoted by I_t. It incurs cost $\ell_{I_t}^{(t)}$. Depending on the feedback model, it gets to known only $\ell_{I_t}^{(t)}$ (partial/bandit feedback) or the entire vector $\ell^{(t)}$ (full/expert feedback).

The external regret of an algorithm on a sequence $\ell^{(1)}, \ldots, \ell^{(T)}$ is defined as

$$\text{Regret}^{(T)} = \mathbf{E}\left[\sum_{t=1}^{T} \ell_{I_t}^{(t)}\right] - \min_i \sum_{t=1}^{T} \ell_i^{(t)} \quad .$$

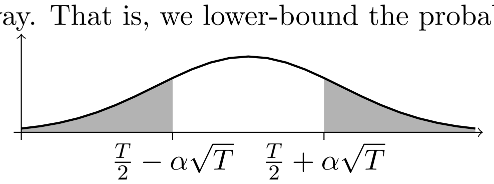
With full feedback, Multiplicative Weights guarantees us that the regret is bounded by $O(\sqrt{T \log n})$ on any sequence of length T with n actions. With partial feedback, Exp3 guarantees regret $O(\sqrt{nT \log n})$.

While the dependence on n is different, the growth in T is the same. Today, we will show that this is actually no surprise. We will prove that for every algorithm there are sequences such that $\text{Regret}^{(T)} = \Omega(\sqrt{T})$, even with only $n = 2$ actions.

1 Anti-Concentration of a Binomial Distribution

In order to derive a lower bound on the regret, we have our algorithm guess coin tosses. We will see that in hindsight it would be much better to always predict heads or tails, depending on the outcomes. For this argument, we will need an "anti-concentration" result on binomial distributions.

Suppose we flip a fair coin T times. How often do we see heads? The expected number of times we see heads is clearly $T/2$. But this does not mean that we always see it exactly $T/2$ times. Our first result shows that there is constant probability that we are at least $\Theta(\sqrt{T})$ away. That is, we lower-bound the probability in the gray area.



Lemma 19.1. *Let $T \in \mathbb{N}$ be even and let X be drawn from a binomial distribution with parameters T and $\frac{1}{2}$. Then for all $\alpha \geq 0$*

$$\mathbf{Pr}\left[\frac{T}{2} - \alpha\sqrt{T} < X < \frac{T}{2} + \alpha\sqrt{T}\right] \leq 2\alpha\frac{\mathrm{e}}{\pi} \quad .$$

Proof. Let

$$J = \left\{ j \in \mathbb{N} \;\middle|\; \frac{T}{2} - \alpha\sqrt{T} < j < \frac{T}{2} + \alpha\sqrt{T} \right\} \quad .$$

We would like to bound the probability of $\mathbf{Pr}[X \in J] = \sum_{j \in J} \mathbf{Pr}[X = j]$. By the definition of the binomial distribution, we have for all $j \in \{0, \ldots, T\}$

$$\mathbf{Pr}[X = j] = \frac{1}{2^T}\binom{T}{j} \quad .$$

We have to bound the binomial coefficient. We can do this using Stirling's approximation, which says

$$\sqrt{2\pi}\ k^{k+\frac{1}{2}}\mathrm{e}^{-k} \leq k! \leq \mathrm{e}\ k^{k+\frac{1}{2}}\mathrm{e}^{-k} \quad \text{for all } k \enspace .$$

This gives us

$$\binom{T}{T/2} = \frac{T!}{(T/2)!} \leq \frac{\mathrm{e}T^{T+\frac{1}{2}}\mathrm{e}^{-T}}{\left(\sqrt{2\pi}(T/2)^{T/2+\frac{1}{2}}\mathrm{e}^{-T/2}\right)^2} = \frac{\mathrm{e}T^{T+\frac{1}{2}}\mathrm{e}^{-T}}{2\pi(T/2)^{T+1}\mathrm{e}^{-T}} = \frac{\mathrm{e}}{\pi}\frac{2^T}{\sqrt{T}} \enspace .$$

Using the monotonicity of binomial coefficients, we have $\binom{T}{j} \leq \frac{\mathrm{e}}{\pi}\frac{2^T}{\sqrt{T}}$ for all j. So we have for all j

$$\mathbf{Pr}\left[X = j\right] = \frac{1}{2^T}\binom{T}{j} \leq \frac{\mathrm{e}}{\pi}\frac{1}{\sqrt{T}}$$

and therefore

$$\mathbf{Pr}\left[X \in J\right] = \sum_{j \in J} \mathbf{Pr}\left[X = j\right] \leq \lvert J \rvert \frac{\mathrm{e}}{\pi}\frac{1}{\sqrt{T}} \enspace .$$

Using that $\lvert J \rvert \leq 2\alpha\sqrt{T}$, the claim follows □

Our next lemma makes use of the observation to show a bound on the expectation.

Lemma 19.2. *Let $T \in \mathbb{N}$ be even and let X be drawn from a binomial distribution with parameters T and $\frac{1}{2}$. Then*

$$\mathbf{E}\left[\min\{X, T - X\}\right] \leq \frac{T}{2} - 0.06\sqrt{T}$$

That is, if we have T fair coin tosses, we consider how often we see heads (X) and how often we see tails ($T - X$). The expectation of the minimum of these two numbers is indeed significantly smaller than $T/2$, although the expectation of each of the numbers is exactly $T/2$.

Proof. Observe that always $\min\{X, T - X\} \leq T/2$. Therefore, for any $\alpha \geq 0$, we can write

$$\begin{aligned}
&\mathbf{E}\left[\min\{X, T - X\}\right] \\
&= \sum_{j=0}^{T/2} j \cdot \mathbf{Pr}\left[\min\{X, T - X\} = j\right] \\
&\leq \left(\frac{T}{2} - \alpha\sqrt{T}\right) \cdot \mathbf{Pr}\left[\min\{X, T - X\} \leq \frac{T}{2} - \alpha\sqrt{T}\right] + \frac{T}{2} \cdot \mathbf{Pr}\left[\min\{X, T - X\} > \frac{T}{2} - \alpha\sqrt{T}\right] \\
&= \frac{T}{2} - \alpha\sqrt{T} + \alpha\sqrt{T} \cdot \mathbf{Pr}\left[\min\{X, T - X\} > \frac{T}{2} - \alpha\sqrt{T}\right] \enspace .
\end{aligned}$$

By Lemma 19.1, we have

$$\mathbf{Pr}\left[\min\{X, T - X\} > \frac{T}{2} - \alpha\sqrt{T}\right] \leq \frac{2\alpha\mathrm{e}}{\pi} \enspace .$$

In combination, this yields

$$\mathbf{E}\left[\min\{X, T - X\}\right] \leq \frac{T}{2} - \alpha\sqrt{T} + \alpha\sqrt{T}\frac{2\alpha\mathrm{e}}{\pi} \enspace .$$

Using, for example, $\alpha = \frac{1}{2}$, we get $\mathbf{E}\left[\min\{X, T - X\}\right] \leq \frac{T}{2} - \frac{1}{2}(1 - \frac{\mathrm{e}}{\pi})\sqrt{T} \leq \frac{T}{2} - 0.06\sqrt{T}$. □

2 Lower Bound on the Regret

Now, we come back to our lower bound on the regret of any algorithm. As mentioned before, we will have our algorithm guess the outcome of fair coin tosses. The driving observation is as follows: Before a coin toss, both outcomes are equally likely. However, in hindsight, either heads or tails will show more often than the other and it would have been good to have predicted this outcome all the time.

Theorem 19.3. *Even for $n = 2$, no algorithm guarantees external regret $o(\sqrt{T})$.*

Proof. Let T be an even square number. We generate a *random* sequence $\ell^{(1)}, \ldots, \ell^{(T)}$. We will show that $\mathbf{E}\left[\mathrm{Regret}^{(T)}\right] \geq 0.06\sqrt{T}$. This also shows that there has to be a sequence of which $\ell^{(1)}, \ldots, \ell^{(T)}$ is this high (akin to the arguments using Yao's principle).

For each t, we set $\ell^{(t)}$ independently to $(1,0)$ or to $(0,1)$ with probability $1/2$ each. That is, we flip a coin and there are two experts: One predicts heads, the other one predicts tails. Observe that in each step, no matter how the algorithm chooses the probabilities, its expected cost will be $1/2$. The more formal reason is as follows. Consider $\ell_{\mathrm{Alg}}^{(t)} = p_1^{(t)} \ell_1^{(t)} + p_2^{(t)} \ell_2^{(t)}$. In this expression, $p_1^{(t)}$ and $p_2^{(t)}$ are random variables. Importantly they are independent of $\ell_1^{(t)}$ and $\ell_2^{(t)}$. Therefore

$$\mathbf{E}\left[\ell_{\mathrm{Alg}}^{(t)}\right] = \mathbf{E}\left[p_1^{(t)}\right] \mathbf{E}\left[\ell_1^{(t)}\right] + \mathbf{E}\left[p_2^{(t)}\right] \mathbf{E}\left[\ell_2^{(t)}\right] = \frac{1}{2}\left(\mathbf{E}\left[p_1^{(t)}\right] + \mathbf{E}\left[p_2^{(t)}\right]\right) = \frac{1}{2} \ .$$

So $\mathbf{E}\left[L_{\mathrm{Alg}}^{(T)}\right] = T/2$, where the expectation is also over the randomization of the sequence.

We have to compare this to $\mathbf{E}\left[\min_i L_i^{(T)}\right]$. Note that $L_1^{(T)}$ and $L_2^{(T)}$ are identically distributed, namely according to a binomial distribution with parameters T and $1/2$. Furthermore $L_1^{(T)} + L_2^{(T)} = T$. Therefore, $\min_i L_i^{(T)} = \min\{L_1^{(T)}, T - L_1^{(T)}\}$. This lets us apply Lemma 19.2, which yields

$$\mathbf{E}\left[\min_i L_i^{(T)}\right] = \mathbf{E}\left[\min\{L_1^{(T)}, T - L_1^{(T)}\}\right] \leq \frac{T}{2} - 0.06\sqrt{T} \ .$$

For the regret, this means

$$\mathbf{E}\left[\mathrm{Regret}^{(T)}\right] = \mathbf{E}\left[L_{\mathrm{Alg}}^{(T)} - \min_i L_i^{(T)}\right] \geq 0.06\sqrt{T} \ .$$

Because this holds in expectation, there has to be a sequence $\ell^{(1)}, \ldots, \ell^{(T)}$ on which the regret is this high. □

3 Comparison to UCB1 and Its Regret Notion

As a quick side remark, we observe that we could have stated the regret-learning questions also in terms of rewards rather than costs. If the adversary fixes reward vectors $g^{(1)}, \ldots, g^{(T)}$ in advance, we can set $\ell_i^{(t)} = 1 - g_i^{(t)}$ and then treat the problem as a cost-minimization problem. For the regret, we then have

$$\mathrm{Regret}^{(t)} = \mathbf{E}\left[\sum_{t=1}^{T} \ell_{I_t}^{(t)}\right] - \min_i \sum_{t=1}^{T} \ell_i^{(t)} = \max_i \sum_{t=1}^{T} g_i^{(t)} - \mathbf{E}\left[\sum_{t=1}^{T} g_{I_t}^{(t)}\right]$$

So, also in terms of the reward, we compare to the best single action in the entire sequence.

This problem is now very similar to the stochastic multi-armed bandits problem, for which we gave the UCB1 algorithm. Recall that there in each step, we choose one of each arms n, each of which has an underlying unknown reward distribution $\mathcal{D}_i$. Our decision is based only on the rewards that we observe. This is in turn depends on which arms we pull.

There are still some subtleties. To capture the stochastic setting, we might assume that $g_i^{(t)}$ is a random variable drawn from distribution $\mathcal{D}_i$. This is indeed a way how the adversary could determine its reward sequence. Recall that UCB1 then gives us a bound on the *expected regret*, which in this notation is

$$\max_i \mathbf{E}\left[\sum_{t=1}^T g_i^{(t)}\right] - \mathbf{E}\left[\sum_{t=1}^T g_{I_t}^{(t)}\right] \quad .$$

Exp3 will give us an even stronger guarantee because on every choice of $g^{(1)}, \ldots, g^{(T)}$, we will get

$$\max_i \sum_{t=1}^T g_i^{(t)} - \mathbf{E}\left[\sum_{t=1}^T g_{I_t}^{(t)}\right] \leq 3\sqrt{nT \ln n} \quad ,$$

where now the expectation is only over the random choices of the algorithm. By taking the expectation over $g^{(1)}, \ldots, g^{(T)}$

$$\mathbf{E}\left[\max_i \sum_{t=1}^T g_i^{(t)}\right] - \mathbf{E}\left[\sum_{t=1}^T g_{I_t}^{(t)}\right] = \mathbf{E}\left[\max_i \sum_{t=1}^T g_i^{(t)} - \sum_{t=1}^T g_{I_t}^{(t)}\right] \leq 3\sqrt{nT \ln n} \quad ,$$

This is a stronger guarantee because $\mathbf{E}\left[\max_i \sum_{t=1}^T g_i^{(t)}\right] \geq \max_i \mathbf{E}\left[\sum_{t=1}^T g_i^{(t)}\right]$. As we have seen in the lower-bound proof today, they can indeed differ by as much as $\sqrt{T}$.

4 Unknown Time Horizon

All these algorithms assumed that we know the time horizon T. Indeed, with a slight modification, they also work for unknown time horizons.

The modified algorithm works as follows. Phase $k \geq 0$ consists of steps $2^k, \ldots, 2^{k+1} - 1$. So it consists of 2^k steps. At the beginning of a phase, we restart the no-regret algorithm with $T' = 2^k$.

Theorem 19.4. *If the algorithm that knows T has regret at most $\alpha\sqrt{T}$ on any sequence of length T then the modified algorithm has regret at most $\frac{\sqrt{2}}{\sqrt{2}-1}\alpha\sqrt{T}$.*

For Multiplicative Weights, we would set $\alpha = 2\sqrt{\ln n}$, for Exp3 $\alpha = 3\sqrt{n \ln n}$. In both cases, we lose only a constant factor in the regret.

Proof. We start $m = \lfloor \log_2 T \rfloor + 1$ phases during T steps. As the last phase might not be complete, we fill up the sequence by $\ell^{(T+1)}, \ldots, \ell^{(2^m-1)}$ with all-zero vectors. This neither changes the cost of a single action nor of the algorithm.

In each phase, we restart the algorithm. Therefore, if P_k are the steps in phase k, we have the regret guarantee

$$\mathbf{E}\left[\sum_{t \in P_k} \ell_{I_t}^{(t)}\right] \leq \min_i \sum_{t \in P_k} \ell_i^{(t)} + \alpha\sqrt{|P_k|} \quad .$$

Now, we take the sum over $k = 0, \ldots, m-1$ on both sides

$$\mathbf{E}\left[\sum_{k=0}^{m-1} \sum_{t \in P_k} \ell_{I_t}^{(t)}\right] \leq \sum_{k=0}^{m-1} \min_i \sum_{t \in P_k} \ell_i^{(t)} + \alpha \sum_{k=0}^{m-1} \sqrt{|P_k|} \ .$$

The first sum, $\mathbf{E}\left[\sum_{k=0}^{m-1} \sum_{t \in P_k} \ell_{I_t}^{(t)}\right] = \mathbf{E}\left[\sum_{t=1}^{T} \ell_{I_t}^{(t)}\right]$ is exactly the expected cost of the algorithm.

For the second sum, we have

$$\sum_{k=0}^{m-1} \min_i \sum_{t \in P_k} \ell_i^{(t)} \leq \min_i \sum_{k=0}^{m-1} \sum_{t \in P_k} \ell_i^{(t)} = \min_i \sum_{t=1}^{T} \ell_i^{(t)} \ .$$

So, the regret is bounded by the third sum

$$\mathrm{Regret}^{(T)} = \mathbf{E}\left[\sum_{t=1}^{T} \ell_{I_t}^{(t)}\right] - \min_i \sum_{t=1}^{T} \ell_i^{(t)} \leq \alpha \sum_{k=0}^{m-1} \sqrt{|P_k|} \ .$$

Using that $|P_k| = 2^k$ gives us

$$\sum_{k=0}^{m-1} \sqrt{|P_k|} = \sum_{k=0}^{m-1} (\sqrt{2})^k = 2\sqrt{\ln n} \frac{(\sqrt{2})^m - 1}{\sqrt{2} - 1} \leq \frac{(\sqrt{2})^m}{\sqrt{2} - 1} \ .$$

Combining this with $(\sqrt{2})^m = 2^{\frac{m}{2}} \leq 2^{\frac{1 + \log_2 T}{2}} = \sqrt{2T}$, we get the regret bound

$$\mathrm{Regret}^{(T)} \leq \alpha \sum_{k=0}^{m-1} \sqrt{|P_k|} \leq \alpha \frac{\sqrt{2T}}{\sqrt{2} - 1} \ . \qquad \square$$