No-Regret Learning: Multi-Armed Bandits 2

Thomas Kesselheim Last Update: June 26, 2020

1 Last Lectures

In the last lecture, we turned the Multiplicative Weights algorithm from the lecture before into one that works with bandit feedback.

We can choose from n actions in every step. An adversary determines the sequence of cost vectors $\ell^{(1)}, \ldots, \ell^{(T)}$ in advance, $\ell_i^{(t)} \in [0,1]$. The sequence is unknown to the algorithm. In step t, the algorithm chooses one of the n actions at random by defining probabilities $p_1^{(t)}, \ldots p_n^{(t)}$. The algorithm's choice in step t is denoted by I_t. The algorithm gets to know $\ell_{I_t}^{(t)}$. The other entries of the cost vector remain unknown.

We used the Multiplicative Weights algorithm in a way that we could reuse the regret bound by computing "fake costs" $\tilde{\ell}_i^{(t)}$. The final combined algorithm then looks as follows, using γ, η, and ρ as parameters.

- Initially, set $w_i^{(1)} = 1$, $p_i^{(1)} = \frac{1}{n}$, for every $i \in [n]$.
- At every time t
  - Define $q_i^{(t)} = (1-\gamma)p_i^{(t)} + \frac{\gamma}{n}$.
  - Choose I_t based on $q^{(t)}$.
  - Define $\tilde{\ell}_{I_t}^{(t)} = \ell_{I_t}^{(t)} / q_{I_t}^{(t)}$ and $\tilde{\ell}_i^{(t)} = 0$ for $i \neq I_t$
  - Multiplicative-Weights Update:
    * Set $w_i^{(t+1)} = w_i^{(t)} \cdot \exp\left(-\eta \frac{1}{\rho} \tilde{\ell}_i^{(t)}\right)$
    * $W^{(t+1)} = \sum_{i=1}^n w_i^{(t+1)}$
    * $p_i^{(t+1)} = w_i^{(t+1)} / W^{(t+1)}$

We set $\gamma = \sqrt[3]{\frac{n \ln n}{T}}$, $\eta = \ln(1-\gamma)$ and $\rho = \frac{n}{\gamma}$ to get a regret bound of $3(n \ln n)^{1/3} T^{2/3}$. Note that we use the weight update $w_i^{(t+1)} = w_i^{(t)} \cdot \exp\left(-\eta \frac{1}{\rho} \tilde{\ell}_i^{(t)}\right)$ instead of $w_i^{(t+1)} = w_i^{(t)} \cdot (1-\eta)^{\frac{1}{\rho}\tilde{\ell}_i^{(t)}}$, which is only a different parameterization.

2 The Exp3 Algorithm

There is a way to improve the regret guarantee to $O(\sqrt{nT \log n})$, which we will get to know today. The algorithm is called Exp3, which stands for "Explore and Exploit with Exponential Weights". And, in fact, we already know the algorithm. It is exactly the one listed above but with a smarter choice of parameters and a more careful analysis.

Our original analysis of the multiplicative-weights update could only deal with cost vectors such that $0 \leq \tilde{\ell}_i^{(t)} \leq \rho$. Now, a single entry $\tilde{\ell}_i^{(t)}$ can be as large as $\frac{n}{\gamma}$. This is why we chose $\rho = \frac{n}{\gamma}$. Exp3 instead sets $\rho = 1$. This means, the update step is much more aggressive than with our previous parameter choice. The vague idea to keep in mind why this is reasonable is that $\tilde{\ell}_i^{(t)} = 0$ most of the time. The fake cost is only non-zero if this is the action that has just been chosen.

The other parameters, γ and η, will be determined later.

3 A Refined Bound of the Multiplicative-Weights Update

The key to prove the regret guarantee of Exp3 is a more careful analysis of the multiplicative-weights update, now allowing $\tilde{\ell}_i^{(t)} > 1$ despite setting $\rho = 1$. We can show the following bound.

Lemma 18.1. *Fix $\tilde{\ell}^{(1)}, \ldots, \tilde{\ell}^{(T)}$ arbitrarily such that $0 \leq \tilde{\ell}_i^{(t)} \leq \frac{1}{\eta}$ for all i and t. Then the vectors $p^{(1)}, \ldots, p^{(T)}$ computed by the multiplicative-weights update (with $\rho = 1$) fulfill*

$$\sum_{t=1}^{T} \sum_{i=1}^{n} p_i^{(t)} \tilde{\ell}_i^{(t)} - \eta \sum_{t=1}^{T} \sum_{i=1}^{n} p_i^{(t)} \left(\tilde{\ell}_i^{(t)} \right)^2 \leq \min_i \sum_{t=1}^{T} \tilde{\ell}_i^{(t)} + \frac{\ln n}{\eta} \quad .$$

Proof. We prove this bound in a very similar way to our original analysis of the multiplicative weights algorithm. We again use the sum of the weights $W^{(t)}$ to (a) lower-bound any expert's cost as well as to (b) upper-bound the algorithm's cost. Combining these two bounds then lets us compare the algorithm's cost to any experts costs.

For part (a), that is the lower bound, we use that for all experts i

$$W^{(T+1)} \geq w_i^{(T+1)} = \exp\left(-\eta \sum_{t=1}^{T} \tilde{\ell}_i^{(t)} \right) \quad .$$

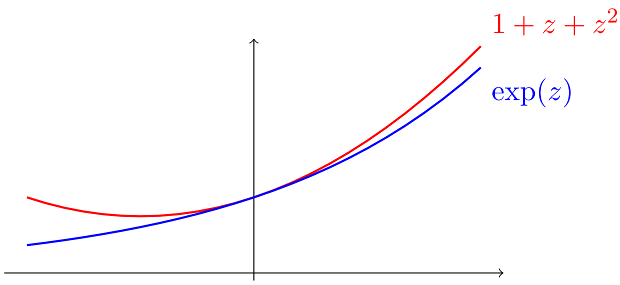
Taking the logarithm, this is equivalent to

$$\ln W^{(T+1)} \geq -\eta \min_i \sum_{t=1}^{T} \tilde{\ell}_i^{(t)} \quad . \tag{1}$$

For part (b), that is the upper bound, we consider the weight changes in step t. We have

$$W^{(t+1)} = \sum_{i=1}^{n} w_i^{(t)} \mathrm{e}^{-\eta \tilde{\ell}_i^{(t)}} \quad .$$

We use that $\mathrm{e}^z \leq 1 + z + z^2$ for $-1 \leq z \leq 1$.



So, we have $\mathrm{e}^{-\eta \tilde{\ell}_i^{(t)}} \leq 1 - \eta \tilde{\ell}_i^{(t)} + \left(\eta \tilde{\ell}_i^{(t)}\right)^2$ because $0 \leq \eta \tilde{\ell}_i^{(t)} \leq 1$. Furthermore, note that we can write $w_i^{(t)} = W^{(t)} p_i^{(t)}$ to get

$$\begin{aligned} W^{(t+1)} &\leq \sum_{i=1}^n w_i^{(t)} \left(1 - \eta \tilde{\ell}_i^{(t)} + \left(\eta \tilde{\ell}_i^{(t)}\right)^2\right) \\ &= \sum_{i=1}^n w_i^{(t)} - \sum_{i=1}^n w_i^{(t)} \eta \tilde{\ell}_i^{(t)} + \sum_{i=1}^n w_i^{(t)} \left(\eta \tilde{\ell}_i^{(t)}\right)^2 \\ &= W^{(t)} \left(1 - \eta \sum_{i=1}^n p_i^{(t)} \tilde{\ell}_i^{(t)} + \eta^2 \sum_{i=1}^n p_i^{(t)} \left(\tilde{\ell}_i^{(t)}\right)^2\right) \enspace . \end{aligned}$$

Repeatedly applying this bound and using that $W^{(1)} = n$, we get

$$W^{(T+1)} \leq n \prod_{t=1}^T \left(1 - \eta \sum_{i=1}^n p_i^{(t)} \tilde{\ell}_i^{(t)} + \eta^2 \sum_{i=1}^n p_i^{(t)} \left(\tilde{\ell}_i^{(t)}\right)^2\right) \enspace .$$

Again, we take the logarithm to get

$$\ln W^{(T+1)} \leq \ln n + \sum_{t=1}^T \ln \left(1 - \eta \sum_{i=1}^n p_i^{(t)} \tilde{\ell}_i^{(t)} + \eta^2 \sum_{i=1}^n p_i^{(t)} \left(\tilde{\ell}_i^{(t)}\right)^2\right) \enspace .$$

We use that $\ln(1+z) \leq z$ for all $z \in \mathbb{R}$ (where defined) to simplify this expression to

$$\ln W^{(T+1)} \leq \ln n - \eta \sum_{t=1}^T \sum_{i=1}^n p_i^{(t)} \tilde{\ell}_i^{(t)} + \eta^2 \sum_{t=1}^T \sum_{i=1}^n p_i^{(t)} \left(\tilde{\ell}_i^{(t)}\right)^2 \enspace . \tag{2}$$

Combining the two bounds on $\ln W^{(T+1)}$, that is, (1) and (2), we get

$$-\eta \min_i \sum_{t=1}^T \tilde{\ell}_i^{(t)} \leq \ln n - \eta \sum_{t=1}^T \sum_{i=1}^n p_i^{(t)} \tilde{\ell}_i^{(t)} + \eta^2 \sum_{t=1}^T \sum_{i=1}^n p_i^{(t)} \left(\tilde{\ell}_i^{(t)}\right)^2 \enspace ,$$

which is equivalent to the claim. □

4 Analysis of Exp3

Based on Lemma 18.1, the remaining analysis of Exp3 works almost the same way as the one for the basic algorithm.

Theorem 18.2. *If $\eta \leq \frac{\gamma}{n}$, Exp3 has expected cost at most*

$$\min_i \sum_{t=1}^T \ell_i^{(t)} + \frac{\ln n}{\eta} + \eta n T + \gamma T \enspace .$$

Proof. Once again, we first fix $I_1, \ldots, I_T$ arbitrarily. This also fixes $\tilde{\ell}^{(1)}, \ldots, \tilde{\ell}^{(T)}$, which are fed into the multiplicative-weights part and this way $p^{(1)}, \ldots, p^{(T)}$ are fixed as well. So, we can invoke Lemma 18.1. Replacing $q_i^{(t)} = (1-\gamma)p_i^{(t)} + \frac{\gamma}{n}$, we have

$$\begin{aligned}\sum_{t=1}^T \sum_{i=1}^n q_i^{(t)} \tilde{\ell}_i^{(t)} &= \sum_{t=1}^T \sum_{i=1}^n \left((1-\gamma)p_i^{(t)} + \frac{\gamma}{n}\right) \tilde{\ell}_i^{(t)} \\ &= (1-\gamma) \sum_{t=1}^T \sum_{i=1}^n p_i^{(t)} \tilde{\ell}_i^{(t)} + \frac{\gamma}{n} \sum_{t=1}^T \sum_{i=1}^n \tilde{\ell}_i^{(t)} \\ &\leq (1-\gamma) \left(\min_i \sum_{t=1}^T \tilde{\ell}_i^{(t)} + \frac{\ln n}{\eta} + \eta \sum_{t=1}^T \sum_{i=1}^n p_i^{(t)} \left(\tilde{\ell}_i^{(t)}\right)^2 \right) + \frac{\gamma}{n} \sum_{t=1}^T \sum_{i=1}^n \tilde{\ell}_i^{(t)} \\ &\leq \min_i \sum_{t=1}^T \tilde{\ell}_i^{(t)} + \frac{\ln n}{\eta} + \eta \sum_{t=1}^T \sum_{i=1}^n q_i^{(t)} \left(\tilde{\ell}_i^{(t)}\right)^2 + \frac{\gamma}{n} \sum_{t=1}^T \sum_{i=1}^n \tilde{\ell}_i^{(t)} \enspace .\end{aligned}$$

Next, we consider how the values $\tilde{\ell}_i^{(t)}$ are derived from the $\ell_i^{(t)}$. To this end, keep $I_1, \ldots, I_{t-1}$ fixed. Like in the analysis of our black-box transformation, we have

$$\mathbf{E}\left[\tilde{\ell}_i^{(t)} \;\middle|\; I_1, \ldots, I_{t-1}\right] = \mathbf{Pr}\left[I_t = i \mid I_1, \ldots, I_{t-1}\right] \cdot \frac{\ell_i^{(t)}}{q_i^{(t)}} + \mathbf{Pr}\left[I_t \neq i\right] \cdot 0 = \ell_i^{(t)} \enspace .$$

So, also

$$\mathbf{E}\left[\sum_{i=1}^n q_i^{(t)} \tilde{\ell}_i^{(t)}\right] = \sum_{i=1}^n \mathbf{E}\left[q_i^{(t)} \tilde{\ell}_i^{(t)}\right] = \sum_{i=1}^n \mathbf{E}\left[q_i^{(t)}\right] \ell_i^{(t)} = \mathbf{E}\left[\ell_{I_t}^{(t)}\right] \enspace .$$

Now, we also have quadratic terms. For these, we can derive

$$\mathbf{E}\left[\left(\tilde{\ell}_i^{(t)}\right)^2 \;\middle|\; I_1, \ldots, I_{t-1}\right] = \mathbf{Pr}\left[I_t = i\right] \cdot \left(\frac{\ell_i^{(t)}}{q_i^{(t)}}\right)^2 + \mathbf{Pr}\left[I_t \neq i\right] \cdot 0 = \frac{\left(\ell_i^{(t)}\right)^2}{q_i^{(t)}} \enspace .$$

This gives us for any choice of $I_1, \ldots, I_{t-1}$

$$\mathbf{E}\left[\sum_{i=1}^n q_i^{(t)} \left(\tilde{\ell}_i^{(t)}\right)^2 \;\middle|\; I_1, \ldots, I_{t-1}\right] = \sum_{i=1}^n q_i^{(t)} \frac{\left(\ell_i^{(t)}\right)^2}{q_i^{(t)}} = \sum_{i=1}^n \left(\ell_i^{(t)}\right)^2 \enspace .$$

As the right-hand side is independent of $I_1, \ldots, I_{t-1}$, this identity also holds for the unconditional expectation

$$\mathbf{E}\left[\sum_{i=1}^n q_i^{(t)} \left(\tilde{\ell}_i^{(t)}\right)^2\right] = \sum_{i=1}^n \left(\ell_i^{(t)}\right)^2 \enspace .$$

Taking the expectation over the bound from the multiplicative weights part, we get

$$\mathbf{E}\left[\sum_{t=1}^T \ell_{I_t}^{(t)}\right] \leq \mathbf{E}\left[\min_i \sum_{t=1}^T \tilde{\ell}_i^{(t)}\right] + \frac{\ln n}{\eta} + \eta \sum_{t=1}^T \sum_{i=1}^n \mathbf{E}\left[q_i^{(t)} \left(\tilde{\ell}_i^{(t)}\right)^2\right] + \frac{\gamma}{n} \sum_{t=1}^T \sum_{i=1}^n \mathbf{E}\left[\tilde{\ell}_i^{(t)}\right]$$

Inserting the above identities, this implies

$$\mathbf{E}\left[\sum_{t=1}^T \ell_{I_t}^{(t)}\right] \leq \min_i \sum_{t=1}^T \ell_i^{(t)} + \frac{\ln n}{\eta} + \eta \sum_{t=1}^T \sum_{i=1}^n \left(\ell_i^{(t)}\right)^2 + \frac{\gamma}{n} \sum_{t=1}^T \sum_{i=1}^n \ell_i^{(t)} \enspace .$$

Finally, we use that $\ell_i^{(t)} \leq 1$ for all i and t. This lets us bound the double sums by nT. (This is not too wasteful because they are multiplied by η or $\frac{\gamma}{n}$, which are small.) Therefore

$$\mathbf{E}\left[\sum_{t=1}^{T} \ell_{I_t}^{(t)}\right] \leq \min_i \sum_{t=1}^{T} \ell_i^{(t)} + \frac{\ln n}{\eta} + \eta n T + \gamma T \ . \qquad \square$$

Corollary 18.3. *Setting* $\eta = \sqrt{\frac{\ln n}{nT}}$, $\gamma = n\eta$, *the external regret of Exp3 is at most* $3\sqrt{nT \ln n}$.

5 Reference

Peter Auer, Nicoló Cesa-Bianchi, Yoav Freund, and Robert E. Schapire. 2003. The Nonstochastic Multiarmed Bandit Problem. SIAM J. Comput. 32, 1 (January 2003), 48-77