E1 245: Online Prediction & Learning **Fall 2015**

Lecture 19 — October 8

Lecturer: Aditya Gopalan *Scribe: Ravi Ranjan*

19.1 Recap

In the last lecture we saw the online mirror descent(OMD) algorithm, in both active as well as lazy versions. We analysed the regret bound for active OMD algorithm. We started studying the problem of 'Online learning with partial information', one version of which is called the Bandits problem. We set up the problem and defined regret at time T,

$$\text{Regret}(T) = \sum_{t=1}^{T} \mathbb{E}\left[l(I_t, t)\right] - \min_{i \in [N]} \sum_{t=1}^{T} l(i, t).$$

The following observations were made:

- Any deterministic algorithm can always be forced to have linear regret, like in the case of 'Online learning with full information'. So, we'll focus on randomized algorithms only.
- Since full information is not available, a prediction-with-experts algorithm like EXP-WTS can't be run 'as-is'.

19.2 Online learning with partial information

Idea: Use a full info strategy, like EXPWTS, with estimates of loss vectors per round. Suppose you want to build an estimate of a vector x in $\mathbb{R}^2$ with sample access to only one coordinate of your choice. The estimate should be unbiased, i.e. $\mathbb{E}[\hat{x}] = x$. One of the strategy can be coin tossing. Toss a coin and estimate as,

$$\hat{x} = \begin{cases} \begin{pmatrix} \frac{x_1}{p} \\ 0 \end{pmatrix}, & \text{if coin} = 0 \\ \begin{pmatrix} 0 \\ \frac{x_2}{1-p} \end{pmatrix}, & \text{if coin} = 1 \end{cases}$$

where,

$$\text{coin} = \begin{cases} 0, & \text{with prob } p \\ 1, & \text{with prob } 1 - p. \end{cases}$$

Here,

$$\mathbb{E}[\hat{x}] = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$

so it is a unbiased estimator.

19.2.1 EXP3 Algorithm

Algorithm 1 EXP3(η)

```
procedure
    Parameter :
        $\eta \geq 0$
    Initialize:
        $p_1 \leftarrow \text{Unif}\{1,2,3,\cdots,N\}$
        $t \leftarrow 1$
    loop:
        Play action $I_t \sim p_t$
        See $l(I_t, p_t)$
        $\tilde{l}(i,t) \leftarrow \begin{cases} l(i,t), & I_t = i \\ 0, & I_t \neq i \end{cases}$
        Update $\forall i$ :
            $p_{t+1}(i) \leftarrow \frac{p_t(i)e^{-\eta\tilde{l}(i,t)}}{\sum_{j=1}^{N} p_t(j)e^{-\eta\tilde{l}(j,t)}}$
```

Here, $l(.) \in [0,1]$ is assumed.

Theorem 19.1. *Regret bound of EXP3(η) ,*

$$\textit{Regret}(T) \leq \frac{\log N}{\eta} + \frac{T\eta N}{2}.$$

Proof: Let,

$$\tilde{l}(i,t) = \frac{l(i,t)}{p_t(i)} \mathbb{1}_{I_t=i} \tag{19.1}$$

and,

$$\tilde{L}_T(i) = \sum_{t=1}^{T} \tilde{l}(i,t).$$

Define the potential function,

$$\phi_t := -\frac{1}{\eta}\log\left(\sum_{i=1}^{N} e^{-\eta\tilde{L}_{t-1}(i)}\right).$$

Now, the per step change in potential, in round t:

$$\begin{aligned}
\phi_{t+1}-\phi_t &= -\frac{1}{\eta}\log\left(\frac{\sum_{i=1}^{N} e^{-\eta\tilde{L}_{t-1}(i)-\eta\tilde{l}(i,t)}}{\sum_{i=1}^{N} e^{-\eta\tilde{L}_{t-1}(i)}}\right)\\
&= -\frac{1}{\eta}\log\left(\sum_{i=1}^{N}\underbrace{\frac{e^{-\eta\tilde{L}_{t-1}(i)}}{\sum_{j=1}^{N} e^{-\eta\tilde{L}_{t-1}(j)}}}_{p_t(i)} e^{-\eta\tilde{l}(i,t)}\right)\\
&= -\frac{1}{\eta}\log\left(\sum_{i=1}^{N} p_t(i)e^{-\eta\tilde{l}(i,t)}\right)\\
&\geq -\frac{1}{\eta}\log\sum_{i=1}^{N} p_t(i)\left(1-\eta\tilde{l}(i,t)+\frac{\left(-\eta\tilde{l}(i,t)\right)^2}{2}\right)\left[\because e^{-x}\leq 1-x+\frac{x^2}{2}\ \forall x\geq 0\right]\\
&= -\frac{1}{\eta}\log\left(1-\eta\sum_{i=1}^{N} l(i,t)\mathbb{1}_{I_t=i}+\frac{\eta^2}{2}\sum_{i=1}^{N}\frac{l^2(i,t)}{p_t(i)}\right)\text{[Putting 19.1]}\\
&\geq -\frac{1}{\eta}\left(-\eta\sum_{i=1}^{N} l(i,t)\mathbb{1}_{I_t=i}+\frac{\eta^2}{2}\sum_{i=1}^{N}\frac{l^2(i,t)}{p_t(i)}\right)[\because \log(1+x)\leq x\ \ \forall x\geq 0]
\end{aligned}$$

Taking conditional expectations on both side,

$$\begin{aligned}
\mathbb{E}\left[\phi_{t+1}-\phi t|H_{t-1}\right] &\geq \mathbb{E}\left[\sum_{i=1}^{N} l(i,t)\mathbb{1}_{I_t=i}|H_{t-1}\right]-\frac{\eta}{2}\mathbb{E}\left[\sum_{i=1}^{N}\frac{l^2(i,t)}{p_t(i)}\mathbb{1}_{I_t=i}|H_{t-1}\right]\\
&= \sum_{i=1}^{N} l(i,t)p_t(i)-\frac{\eta}{2}\sum_{i=1}^{N} l^2(i,t)\\
&\qquad\qquad \left[\text{Exchanging summation and expectation, and putting}\mathbb{E}\left[\mathbb{1}_{I_t=i}|H_{t-1}\right]=p_t(i)\right]\\
&\geq \sum_{i=1}^{N} l(i,t)p_t(i)-\frac{\eta}{2}N \qquad [\because l^2(i,t)\leq 1].
\end{aligned}$$

Again applying expectations on both side,

$$\mathbb{E}\left[\mathbb{E}\left[\phi_{t+1}-\phi_t|H_{t-1}\right]\right]\geq\mathbb{E}\left[\sum_{i=1}^{N} l(i,t)p_t(i)\right]-\frac{\eta}{2}N.$$

Applying law of iterated expectation on LHS,

$$\mathbb{E}\left[\phi_{t+1}-\phi_t\right] \geq \mathbb{E}\left[l(I_t,t)\right]-\frac{\eta}{2}N.$$

Summing it over $t=1,2,\cdots,T$

$$\mathbb{E}\left[\phi_{T+1}\right]-\mathbb{E}\left[\phi_1\right] \geq \mathbb{E}\left[\sum_{t=1}^{T} l(I_t,t)\right]-\frac{T\eta}{2}N. \tag{19.2}$$

On other hand,

$$\begin{aligned}\phi_{T+1}-\phi_1 &= -\frac{1}{\eta}\log\left[\frac{\sum_{i=1}^{N} e^{-\eta\tilde{L}_T(i)}}{N}\right]\\ &\leq -\frac{1}{\eta}\log\frac{e^{-\eta\tilde{L}_T(i^*)}}{N} \qquad\qquad \left[\text{where}, i^* = \arg\min_{i\in[N]}\tilde{L}_T(i)\right]\\ &\leq \tilde{L}_T(i^*)+\frac{1}{\eta}\log N\end{aligned}$$

Taking expectation both side,

$$\mathbb{E}\left[\phi_{T+1}\right]-\mathbb{E}\left[\phi_1\right] \leq L_T(i^*)+\frac{1}{\eta}\log N \tag{19.3}$$

Putting together 19.2 and 19.3,

$$\mathbb{E}\left[\sum_{t=1}^{T} l(I_t,t)\right]-L_T(i^*) \leq \frac{1}{\eta}\log N+\frac{T\eta N}{2}. \tag{19.4}$$

□

Note

- The optimal regret of EXP3 is $O(\sqrt{TN\log N})$. If compared with the EXPWTS algorithm, where optimal regret is $O(\sqrt{T\log N})$, the regret of EXP3 is $\sqrt{N}$ times, which is the price for not knowing full information.
- Here,

$$\text{var}[\tilde{l}(i,t)] = \text{var}\left[\frac{l(i,t)}{p_t(i)}\mathbb{1}_{I_t-i}\right] = \left(\frac{l(i,t)}{p_t(i)}\right)^2 p_t(i)(1-p_t(i)) = \frac{l^2(i,t)}{p_t(i)}(1-p_t(i)).$$

$$\implies \text{var}[\tilde{l}(i,t)] \xrightarrow{p_i(t)\to 0} \infty.$$

So, $\text{var}[\tilde{l}(i,t)]$ can be very large implying that the tail of the regret could be very heavy, and as a consequence tail bounds on regret are not possible.

- The minimax-optimal regret is $O(\sqrt{TN})$. If logarithmic factor $\sqrt{\log N}$ is disregarded, then EXP3 is essentially minimax-optimal. The bond $(\sqrt{TN \log N})$ turns out to be minimax optimal. Also, it is shown that INF(Implicitly Normalized Forecaster) algorithm actually achieves the minimax-optimal[3].

Bibliography

[1] Gábor Bartók, Dávid Pál, Csaba Szepesvári and István Szita *Online learning - CMPUT 654, Lecture Notes, October 2011*
http://david.palenica.com/papers/online-learning-lecture-notes/online-learning-lecture-notes-2011-Oct-20.pdf

[2] Jacob Abernethy, *Prediction and Learning: It's Only a Game*, Scibed Lecture Notes, Fall 2013
http://web.eecs.umich.edu/ jabernet/eecs598course/fall2013/web/notes/lec20_111313.pdf

[3] Jean-Yves Audibert, Sébastien Bubeck, *Minimax Policies for Bandits Games*, Submitted to *Journal of Machine Learning Research*
http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.210.2852&rep=rep1&type=pdf