IE613: Online Machine Learning **Jan-Apr 2016**

Lecture 8: Adversarial Bandits: EXP3 Algorithm

Lecturer: M. K. Hanawal *Scribes: Saumya Suri*

8.1 Adversarial Bandit Setting

The hypothesis we used earlier will now be called Actions/Arms. Let there be K actions and let in each round the player selects one out of the K actions and incurs a loss. Unlike the full information setting, where the player had information about the losses due to every possible hypothesis, in the Adversarial Bandit setting,the player has the information of only the action he played in a round. Thus in each round he gets only $1/K$ of the full information. Hence we can say that T rounds in a full information setting is equivalent to KT rounds of the Adversarial bandit setting.

Let $I_t \in [K]$ be the arm played by the player in round t and $l_{i,t}$ be the loss of arm i in round t. Thus, at each time step $t = 1, 2, ...$, the adversary assigns to each arm $i = 1, ..., K$ a loss $l_{i,t}$. We assume that the losses are bounded such that $l_{i,t} \in [0, 1] \; \forall \; t$.

Algorithm 1 Template of the adversarial bandit setting

1: **for** $t = 1, 2, \ldots, T$ **do**
2: Adversary selects $l_{i,t}$, $i \in [K]$
3: Players selects an arm $I_t \in [K]$
4: Player observes only $l_{I_t,t}$ i.e, loss corresponding to only that arm

For example,
If suppose the adversary has assigned losses $l_{1,t}, l_{2,t}, ..., l_{10,t}$ for 10 actions in round t and let say the player plays action 5 in round t, then he would be able to observe only $l_{5,t}$ in that round.

Regret of the algorithm that picks I_t in round t is:

$$\widetilde{R}_T = \sum_{t=1}^{T} l_{I_t,t} - \min_{k=1,2...K} \sum_{t=1}^{T} l_{k,t} \tag{8.1}$$

where, $\sum_{t=1}^{T} l_{I_t,t}$ is the total loss incurred by the player for playing I_t in T rounds and $\min_{k=1,2...K} \sum_{t=1}^{T} l_{k,t}$ is the minimum loss he would have incurred for playing one particular arm for T rounds.

The randomness of the losses and the arm chosen by the player makes $\widetilde{R}_T$ a random quantity.
Taking Expectation on the sides in (8.1), we get,

$$\begin{aligned} R_T = E(\widetilde{R}_T) = & \quad E\Big(\sum_{t=1}^{T} l_{I_t,t}\Big) - E\Big(\min_{k=1,2\ldots K} \sum_{t=1}^{T} l_{k,t}\Big) \\ \geq & \quad E\Big(\sum_{t=1}^{T} l_{I_t,t}\Big) - \min_{k=1,2\ldots K} E\Big(\sum_{t=1}^{T} l_{k,t}\Big) \end{aligned}$$

The term $E\Big(\sum_{t=1}^{T} l_{I_t,t}\Big) - \min_{k=1,2\ldots K} E\Big(\sum_{t=1}^{T} l_{k,t}\Big)$ is known as **Pseudo Regret** and is denoted by $\overline{R}_T$. Thus we have,

$$R_T \geq \overline{R}_T$$

We use Pseudo regret instead of the usual Expected regret(R_T) since $E\Big(\min_{k=1,2\ldots K} \sum_{t=1}^{T} l_{k,t}\Big)$ is difficult to compute as compared to $\min_{k=1,2\ldots K} E\Big(\sum_{t=1}^{T} l_{k,t}\Big)$.

Oblivious Adversary: If before setting the loss value, the adversary does not know which arm that will be chosen by the player then the adversary is known as Oblivious adversary For oblivious adversary, $R_T = \overline{R}_T$

8.1.1 Pseudo Regret bounds

EXP3 (Exponential weights for Exploration and Exploitation) Algorithm

The parameters for the Exp3 algorithm are,

- A non increasing sequence of real numbers $(n_t)_{t\in N}$
- P_1 , a uniform distribution over [K]

Algorithm 2 EXP3 Algorithm

Input: A non increasing sequence of real numbers $(n_t)_{t\in N}$
for $t = 1, 2, \ldots, T$ **do**
  Draw an arm I_t from the probability distribution P_t
  Players selects an arm $I_t \in [K]$
  for Each arm $i = 1, 2, ...K$ **do**
    compute the estimated loss, $\widetilde{l}_{i,t} = \frac{l_{i,t}}{P_{i,t}} \mathbb{1}_{\{I_t=i\}}$
    update the cumulative loss, $\widetilde{L}_{i,t} = \widetilde{L}_{i,t-1} + \widetilde{l}_{i,t}$
    Compute the new probability distribution:
    $P_{t+1} = \big(P_{1,t+1}, P_{2,t+1}.....P_{K,t+1}\big)$ where $P_{i,t+1} = \frac{exp(\eta_t \widetilde{L}_{i,t})}{\sum_{k=1\ldots K} \exp -\eta_t \widetilde{L}_{k,t}}$

The player starts with a uniform distribution P_1 and then in each round pick an arm according to the distribution P_t. We let that arm be I_t and for all arms, calculate the estimated loss,

$$\widetilde{l}_{i,t} = \frac{l_{i,t}}{P_{i,t}} \ \mathbb{1}_{\{I_t=i\}} \tag{8.2}$$

where, $P_{i,t}$ is the probability of choosing arm i in round t.

We can further show that $\widetilde{l}_{i,t}$ is an unbiased Estimator (w.r.t I_t) of $l_{i,t}$ i.e,

$$E(\widetilde{l}_{i,t}) = l_{i,t}$$

since,

$$\sum_{k=1,\dots K} P_{k,t} \frac{l_{k,t}}{P_{k,t}} \mathbb{1}_{\{k=i\}} = \sum_{k=1,\dots K} l_{k,t} \mathbb{1}_{\{k=i\}} = l_{i,t}$$

Thus, though only the loss of the played arm is observed, we build an unbiased estimator for the loss of any other arm as well.

Theorem 8.1 ***Pseudo Regret of Exp3****:*

If Exp3 is run with $\eta_t = \eta = \sqrt{\frac{2 \ln K}{TK}}$ *then the Pseudo regret, i.e,* $\overline{R}_T$ *is bounded above by* $\sqrt{2KT \ln K}$*, i.e,*

$$\overline{R}_T \leq \sqrt{2KT \ln K}$$

Moreover, if Exp3 is run with $\eta_t = \eta = \sqrt{\frac{\ln K}{tK}}$ *then then the Pseudo regret, i.e,* $\overline{R}_T$ *is bounded above by* $2\sqrt{KT \ln K}$ *i.e,*

$$\overline{R}_T \leq 2\sqrt{KT \ln K}$$