EE6106 Lecture 4 (Date: 23 January 2024)

Adversarial Bandits

The information table can be written as an array of the form

$$y = \begin{bmatrix} y_{1,1} & --- & y_{1,T} \\ \vdots & y_{i,T} & \\ y_{k,1} & \cdots & y_{k,T} \end{bmatrix}$$

we are not assuming any kind of underlying distribution for the losses

loss of expert /arm i at time T

Randomization is necessary in the presence of an adversary.
If the environment is however, structured we can have deterministic algos as well.

Protocol

1. At time $t \geq 1$,
   - algo picks arm $A_t \in [K]$
   - loss $Y_{A_t,t}$ is revealed

Goal

Minimize $\mathbb{E}\left[\sum_{1}^{T} Y_{A_t,t}\right] - \left(\min_i \sum_{i}^{T} y_{it}\right)$ ← try and make this sublinear

$$\underbrace{\mathbb{E}\left[\sum_{1}^{T} Y_{A_t,t}\right] - \left(\min_i \sum_{i}^{T} y_{it}\right)}$$

Let's say start calling this to be "$\cancel{\mathbb{E}} R_T$" instead of $\mathbb{E}[R_T]$

Idea : Replace $Y_{j,t}$ by estimate $\hat{Y}_{j,t}$, which is an unbiased estimator
↓
(conditionally unbiased)

Say p_{jt} = probability of picking arm j at time t given I_{t-1}

$$I_{t-1} = (A_1, Y_1, A_2, Y_2, \dots, A_{t-1}, Y_{t-1})$$

Y_t : observed loss, and hence $Y_t = Y_{A_t,t}$

$$\hat{Y}_{j,t} = \frac{\mathbb{1}_{\{A_t=j\}} Y_t}{p_{j,t}}$$

indicator of subsequently actually picking action j

$$= \begin{cases} 0 & i \neq A_t \\ \dfrac{Y_{jt}}{p_{jt}} & j = A_t \end{cases}$$

We know that $\mathbb{E}[\hat{Y}_{j,t} \mid I_{t-1}] = Y_{jt}$

The algorithm does not always have access to Y_{jt} so we can instead use $\hat{Y_{j,t}}$.

$$Var(\hat{Y_{jt}} \mid I_{t-1}) = \frac{Y_{jt}^2 (1-p_{jt})}{p_{jt}}$$

If $p_{jt} \to 0$, then the variance tends to be QUITE LARGE

$$\hat{L_{j,t}} \doteq \sum_{1}^{T} \hat{Y_{j,t}}$$

EXP3 (EXPONENTIAL ALGORITHM FOR EXPLORE & EXPLOIT)

(There is also an EXP4 algorithm, lol)

At time $t \geq 1$,

1) $$p_{jt} = \frac{e^{-\eta \hat{L}_{j,t-1}}}{\sum_i e^{-\eta \hat{L}_{i,t-1}}}$$

2) Pick arm $A_t \sim (p_{jt})_{j=1}^{K}$.

3) $\hat{L_{jt}} = \hat{L_{j,t-1}} + \hat{Y_{j,t}}$ ... (update step)

If we replace $\hat{Y}$ by Y, then this algorithm is exactly similar to REWMA

THEOREM For Exp 3,

$$R_T \leq \frac{\log K}{\eta} + \eta TK$$

$$\text{with } \eta = \sqrt{\frac{\log K}{TK}}, \quad R_T \leq 2\sqrt{TK \log K}$$

PROOF : $R_{i,T} = \mathbb{E}\left[\sum_{t=1}^{T} Y_t\right] - \sum_{i}^{T} Y_{i,t}$ | Expected loss of algorithm relative to an arbitrarily selected arm

Step 1 $\hat{L_{i,t}} = \sum_{s=1}^{t} \hat{Y_{i,s}} \longrightarrow$ estimate of loss of algorithm ~~upto~~ of arm i

$$\mathbb{E}[\hat{L_{i,t}}] = \sum_{s=1}^{T} \mathbb{E}[\cancel{\mathbb{E}_{i,t}}\ \mathbb{E}[\hat{Y_{i,s}} \mid I_{s-1}]]$$

$$\boxed{\therefore \mathbb{E}[\hat{L_{i,t}}] = \sum_{1}^{T} y_{i,s}} \quad \cdots\cdots (1)$$

$$\mathbb{E}[y_t \mid I_{t-1}] = \sum_{j=1}^{k} p_{jt} y_{jt}$$

$$\hat{L}_T = \sum_{t=1}^{T} \sum_{j=1}^{K} p_{j,t} \hat{y}_{j,t}$$

$$\mathbb{E}[\hat{L}_T] = \sum_{t=1}^{T} \sum_{j=1}^{K} \mathbb{E}[\mathbb{E}[p_{jt} \hat{y}_{jt} \mid I_{t-1}]]$$

$$= \sum_{t=1}^{T} \sum_{1}^{K} \mathbb{E}[p_{jt} y_{jt}]$$

$$= \sum_{1}^{T} \mathbb{E}[y_t]$$

We shall then write

$$R_{i,T} = \mathbb{E}[\hat{L}_T] - \mathbb{E}[\hat{L_{i,T}}]$$

$\longrightarrow$ regret of algo rel to arm is NOT regret of arm.

Step 2

$$W_t = e^{\eta t} \sum_{j=1}^{K} e^{-\eta \hat{L}_{j,t}}$$

Lower Bound : $W_T \geq e^{\eta T} e^{-\eta \hat{L}_{i,T}}$

Upper Bound : $$\frac{W_t}{W_{t-1}} = e^{\eta} \left(\frac{\sum_{j=1}^{K} e^{-\eta \hat{L}_{j,t-1}} e^{-\eta \hat{y}_{j,t}}}{\sum_{j=1}^{K} e^{-\eta \hat{L}_{j,t-1}}} \right)$$

$\rightarrow$ probability of choosing arm

$$= e^{-\eta} \sum p_{jt} e^{-\eta \hat{y}_{j,t}}$$

$$= \sum p_{jt} e^{\eta(1 - \hat{y}_{jt})}$$

For $x \leq 1$,

$$1 + x \leq e^x \leq 1 + x + x^2$$

$$\therefore \frac{W_t}{W_{t-1}} \le \sum p_{jt}\left(1+\eta(1-\hat{Y}_{jt}) + \eta^2(1-\hat{Y}_{jt})^2\right)$$

$$= 1 + \eta \sum p_{jt}(1-\hat{Y}_{jt}) + \eta^2 \sum p_{jt}(1-\hat{Y}_{jt})^2$$

$$\frac{W_t}{W_{t-1}} \le e^{\eta \sum p_{jt}(1-\hat{Y}_{jt}) + \eta^2 \sum p_{jt}(1-\hat{Y}_{jt})^2}$$

summation is wrt
the index of arms $\{1, \dots, k\}$

$$\frac{W_T}{K} = \prod_{1}^{T} \frac{W_t}{W_{t-1}} \le \underbrace{e^{\eta \sum\sum p_{jt}(1-\hat{Y}_{jt}) + \eta^2 \sum\sum p_{jt}(1-\hat{Y}_{jt})^2}}_{\text{call this } \mathbb{N}}$$

Combining the LB & UB, we have

$$e^{\eta T} e^{-\eta L_{i,T}} \le K\mathbb{N}$$

$\sum\sum p_{jt}$ $\quad$ $\sum\sum p_{jt} Y_{jt}$

Taking logarithms

$$\eta T - \eta \hat{L}_{i,T} \le \log K + \eta(T - \hat{L}_T) + \eta^2 \sum\sum p_{jt}(1-\hat{Y}_{jt})^2$$

$$\therefore T - \hat{L}_{i,T} \le \frac{\log K}{\eta} + T - \hat{L}_T + \eta \sum_{t=1}^{T}\sum_{j=1}^{k} p_{jt}(1-\hat{Y}_{jt})^2$$

<u>Note</u>: $\hat{Y}_{jt}$ can be arbitrarily large though $Y_{jt} \in [0,1]$
we wish to bound the expectation on the RHS

<u>Lemma</u>: $\mathbb{E}\left[\sum\sum p_{jt}(1-\hat{Y}_{jt})^2\right] \le kT$

$$\therefore \hat{L}_T - \hat{L}_{i,T} \le \frac{\log K}{\eta} + \eta \sum_{t=1}^{T}\sum_{j=1}^{K} p_{jk}(1-\hat{Y}_{jt})^2$$

<u>Proof of lemma</u>:

$$\mathbb{E}\left[\sum\sum p_{jt}(1-\hat{Y}_{jt})^2\right] = \mathbb{E}\left[\sum_{1}^{T}\sum_{1}^{k} p_{jt}(1+\hat{Y}_{jt}^2 - 2\hat{Y}_{jt})\right]$$

$$\mathbb{E}\left[\sum\sum p_{jt}(1-\hat{y}_{jt})^2\right] = \sum_{t=1}^{T} \mathbb{E}\left[1-2y_t+\sum p_{jt}\frac{\mathbb{1}\{A_t=j\}\, y_{jt}^2}{p_{jt}^2}\right]$$

$$= \sum_{1}^{T} \mathbb{E}\left[1-2y_t+\sum_j y_{jt}^2\right]$$

because

$$\mathbb{E}\left[\sum p_{jt}\, \mathbb{1}_{\{A_t=j\}} \ldots\right] = \mathbb{E}(y_{jt} \ldots$$

$$= \sum_{1}^{T} \mathbb{E}\Big[\underbrace{(1-y_t)^2}_{\leq 1} + \underbrace{\sum_{j \neq A_t} y_{jt}^2}_{\leq (k-1)}\Big]$$

Just completing squares

random since A_t is random

$$\leq kT$$

Hence $R_T \leq \frac{\log K}{\eta} + \eta TK$.

<u>Note</u>: A subtle difference in the bounds between REWMA & EXP3 is that there is a factor of K appearing here. Thus, the regret is of the order $O(\sqrt{K})$ worse as compared to the full information setting.

choose $\eta = \sqrt{\frac{\log K}{TK}}$, we get $\longrightarrow$ $R_T \leq 2\sqrt{TK\log K}$

The actual payment comes from here

① $\mathbb{E}\left[\sum\sum p_{jt}(1-\hat{y}_{jt})^2\right] \leq kT$

② However, if we replace $\hat{y_{jt}}$ by y_{jt}, so as to get to the full information setting, the bound we would have got is

$$\mathbb{E}\Big[\sum\sum p_{jt}\underbrace{(1-y_{jt})^2}_{\leq 1}\Big] \leq T$$

<u>Note</u>: If the losses are unbounded, the nature of bounds obtained are much much weaker

<u>Ref</u>: Lattimore presents identical proofs

(From now on, we shall be following Lattimore)

M T W T F S S
Page No.:
Date:
YOUVA

The high probability bounds for the regrets obtained in this algorithm are quite bad. There exists a refined algorithm for this called the Exp3 - IX