18/4/20

$$\hat{X}_{ti} = \left[\frac{\mathbb{1}\{A_t = i\}\, X_t}{P_{ti}}\right]$$

1

$\frac{x_{ti}}{\cancel{P_{ti}}}$

→ 0

Analysis of EXP-3:

Recall $\hat{X}_{ti} = 1 - \frac{\mathbb{1}(A_t = i)}{P_{ti}}(1 - X_t)$

$$\mathbb{E}[\hat{X}_{ti}] : P_{ti} - (1 - X_t) + (1 - P_{ti}) \cdot 1 = \cancel{P_{ti}} - \cancel{1} + X_t + \cancel{1} - \cancel{P_{ti}} = X_t$$

Alg:

1) Input n, k, η
2) $\hat{S}_{0i} = 0 \;\forall i$
3) **for** $t = 1, \dots, n$ **do**
4) Calculate $P_{ti} = \dfrac{\exp(\eta \hat{S}_{t-1,i})}{\sum_{j=1}^{k} \exp(\eta \hat{S}_{t-1,j})} \leftarrow W_{t-1}$
5) Sample $A_t \sim P_t$ & observe X_t
6) $\hat{S}_{ti} = \hat{S}_{t-1,i} + 1 - \dfrac{\mathbb{1}\{A_t = i\}(1 - X_t)}{P_{ti}}$
7) end for

$$\mathbb{E}(\hat{S}_{ti}) = \cancel{P_{ti}}\hat{S}_{t-1,i} + \cancel{P_{ti}} - (1 - X_{ti}) + (1 - \cancel{P_{ti}})\hat{S}_{t-1,i} + 1 - \cancel{P_{ti}} = \hat{S}_{t-1,i} + \underbrace{X_{ti}}$$

- So, if arm j is pulled;

$$\hat{S}_{tj} = \hat{S}_{t-1,j} + 1 - \frac{(1 - \overset{x_{tj}}{X_t})}{P_{tj}}$$

- If arm j is not pulled.

$$\hat{S}_{tj} = \hat{S}_{t-1,j} + 1$$

$$\boxed{E[\hat{S}_{tj}] = \hat{S}_{t-1,j} + x_{tj}}$$

<u>Thm:</u> Let $x \in [0,1]^{n \times k}$ & π be the policy EXP-3 with $\eta = \sqrt{\log(k)/nk}$; then

$$\underbrace{R_n(\pi, x) \leq 2\sqrt{nk \log k}}$$

<u>Proof!</u> Define $R_{ni} = \sum_{t=1}^{n} x_{ti} - E\left[\sum_{t=1}^{n} X_t\right]$

Using arm i always

- We will bound R_{ni} for all i.

<u>Let i be fixed:</u>

$$\mathbb{E}\big[\underbrace{\hat{S}_{ni}}\big] = \mathbb{E}\left[\sum_{t=1}^{n} \hat{X}_{ti}\right]$$

$$= \sum_{t=1}^{n} \underbrace{\mathbb{E}\left[\hat{X}_{ti}\right]}$$

$$= \sum_{t=1}^{n} \left\{ (1-P_{ti}) \cdot 1 + P_{ti}\left[1 - \frac{(1-x_{ti})}{P_{ti}}\right]\right.$$

$$= \sum_{t=1}^{n} 1 - \cancel{P_{ti}} + \cancel{P_{ti}} - (1-x_{ti})$$

$$\mathbb{E}\left[\hat{S}_{ni}\right] = \sum_{t=1}^{n} x_{ti} \qquad \text{(unbiased estimator)}$$

$$\mathbb{E}_t\left[X_t\right] = \sum_{i=1}^{k} P_{ti} \, x_{ti} = \sum P_{ti} \, \mathbb{E}_t\left[\hat{X}_{ti}\right],$$

<u>Now</u> $\mathbb{E}\left[\mathbb{E}_t\left[X_t\right]\right] = \mathbb{E}\left[X_t\right]$.

$$\therefore \quad R_{ni} = \mathbb{E}\left[\hat{S}_{ni}\right] - \mathbb{E}\Big[\overbrace{\sum P_{ti} \, \mathbb{E}_t\left[\hat{x}_{ti}\right]}^{\mathbb{E}\left[\mathbb{E}\left[X_t\right]\right]}\Big]$$

$$= \mathbb{E}\left[\hat{S}_{ni} - \sum_{t=1}^{n}\sum_{i=1}^{k} p_{ti}\hat{x}_{ti}\right]$$

Define: $\hat{S}_n = \sum_{t=1}^{n}\sum_{i=1}^{n} p_{ti}\hat{x}_{ti}$

$$\therefore R_{ni} = \mathbb{E}\left[\hat{S}_{ni} - \hat{S}_n\right]$$

we want to bound: $\exp(\eta \hat{S}_{ni})$

$$\exp(\eta \hat{S}_{ni}) \leq \sum_{j=1}^{k} \exp(\eta \hat{S}_{nj})$$

Set $W_t = \sum_{j=1}^{k} \exp(\eta \hat{S}_{tj})$

$$P_{tj} = \frac{\exp(\eta \hat{S}_{tj})}{W_t}$$

$$\therefore \exp(\eta \hat{S}_{ni}) \leq W_n = W_0 \cdot \frac{W_1}{W_0} \frac{W_2}{W_1} \cdots \frac{W_n}{W_{n-1}}$$

$$= k \frac{W_1}{W_0} \frac{W_2}{W_1} \cdots \frac{W_n}{W_{n-1}}$$

$\exp(\cdot) \exp(\cdot)$

$$= \exp(\Sigma$$

$$\frac{W_t}{W_{t-1}} = \frac{\sum_j \exp(\eta \hat{S}_{tj})}{W_{t-1}} = \frac{\sum_j \exp(\eta(\hat{S}_{t-1,j} + \hat{X}_{tj}))}{W_{t-1}}$$

$$= \sum_{j=1}^{k} \underbrace{\frac{\exp(\eta \hat{S}_{t-1,j})}{W_{t-1}}} \underbrace{\exp(\eta \hat{X}_{tj})}$$

$$= \sum_{j=1}^{k} \underbrace{P_{tj}} \underbrace{\exp(\eta \hat{X}_{tj})} \leftarrow e^{\eta \hat{X}_{tj}}$$

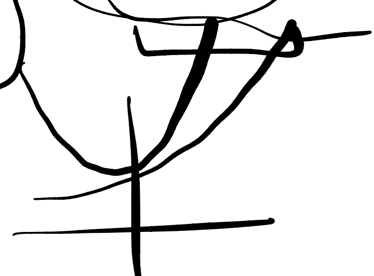
Now $\hat{X}_{tj} \le 1 \quad \forall t, j$ $\qquad 1 - \frac{\{A_t = j\}(1 - X_t)}{P_{tj}}$ $\quad \in [0,1]$

[This would not be true if we used

$$\hat{X}_{tj} = \frac{\mathbb{1}\{A_t = j\} X_t}{P_{tj}}$$
]

$$\boxed{\exp(x) \le 1 + x + x^2 \quad \forall x \le 1.}$$

$$\exp(x) = 1 + x + \frac{x^2}{2!} + \frac{x^3}{2!} + \cdots$$



$$\exp(x) \le 1 + x + \frac{x^2}{2!} \quad \forall x \le 0$$

$$\therefore \frac{W_t}{W_{t-1}} \leq \sum P_{tj}\left(1+\eta \hat{x}_{tj} + \eta^2 \hat{x}_{tj}^2\right)$$

$$\leq 1 + \sum \eta P_{tj}\hat{x}_{tj} + \eta^2 \hat{x}_{tj}^2 P_{tj}$$

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

$$\therefore \exp(\eta \hat{S}_{ni}) \leq \prod_{t=1}^{n} \exp\left(\sum_{j=1}^{k} \eta P_{tj}\hat{x}_{tj} + \eta^2 \sum_{j=1} \hat{x}_{tj}^2 P_{tj}\right)$$

$$\leq k \exp\left(\eta \sum_{t=1}^{n} \sum_{j=1}^{k} P_{tj}\hat{x}_{tj} + \eta^2 \sum_{j=1}^{k} \sum_{t=1}^{n} \hat{x}_{tj}^2 P_{tj}\right)$$

$$\leq k \exp\left(\eta \hat{S}_n + \eta^2 \sum_{j=1}^{k} \sum_{t=1}^{n} \hat{x}_{tj}^2 P_{tj}\right)$$

$$\therefore \eta \hat{S}_{ni} \leq \log k + \eta \hat{S}_n + \eta^2 \sum_{t=1}^{n} \sum_{j=1}^{k} \hat{x}_{tj}^2 P_{tj}$$

$$\therefore \hat{S}_{ni} - \hat{S}_n \leq \frac{\log k}{\eta} + \eta \sum_{t=1}^{n} \sum_{j=1}^{k} \hat{x}_{tj}^2 P_{tj}$$

Now $\mathbb{E}\left[\hat{S}_{ni} - \hat{S}_n\right] = R_{ni}$.

$$\mathbb{E}\left[\hat{S}_{ni} - \hat{S}_n\right] \le \frac{\log k}{\eta} + \mathbb{E}\left(\sum_{t=1}^{n}\sum_{j=1}^{k} P_{tj}\,\hat{X}_{tj}^2\right)$$

$$\mathbb{E}\left[\sum_{t=1}^{n}\sum_{j=1}^{k} P_{tj}\Big(1 - \underbrace{\frac{\mathbb{1}(A_t=j)\,y_{tj}}{P_{tj}}}_{\hat{Y}_{tj}}\Big)^2\right]$$

$y_{tj} = (1 - x_{tj})$

$$= \sum_{t=1}^{n}\mathbb{E}\left(\sum_{j=1}^{k} P_{tj}\left(1 - 2\,\frac{\mathbb{1}\{A_t=j\}\,y_{tj}}{P_{tj}} + \frac{\mathbb{1}\{A_t=j\}\,y_{tj}^2}{P_{tj}^2}\right)\right)$$

$$= \sum_{t=1}^{n}\mathbb{E}\left(1 - 2 Y_t + \mathbb{E}_t\left(\sum_{j=1}^{k}\frac{\mathbb{1}\{A_t=j\}\,y_{tj}^2}{P_{tj}}\right)\right)$$

$$= \sum_{t=1}^{n} \mathbb{E}\left[1 - 2y_t + \sum_{j=1}^{k} y_{tj}^2\right]$$

$$= \sum_{t=1}^{n} \mathbb{E}\left[(1-y_t)^2 + \sum_{j \neq A_t} y_{tj}^2\right]$$

$$\leq \sum_{t=1}^{n} \mathbb{E}\left(1 + (k-1)\right)$$

$$\leq \underline{nk}.$$

$$\therefore \quad R_{ni} \leq \frac{\log(k)}{\eta} + \eta n k.$$

Setting $\eta = \sqrt{\dfrac{\log(k)}{nk}}$

$$\leq 2\sqrt{nk\log k}$$

$$(1-\hat{x}_{ti}) = \frac{\mathbb{1}\{A_t = i\}(1-x_t)}{P_{ti}}$$

$$P_{ti}(1-\hat{x}_{ti}) = \mathbb{1}\{A_t = i\}(1-x_t)$$

$$\therefore \sum P_{ti}(1-\hat{x}_{ti}) = \sum_i \mathbb{1}\{A_t = i\}(1-x_t)$$

$$\therefore 1 - \sum P_{ti}\hat{x}_{ti} = (1-x_t)$$

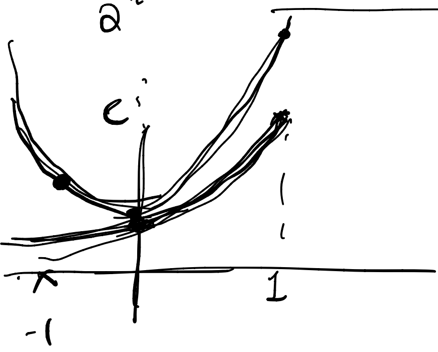
$$\therefore \underline{x_t = \sum P_{ti}\hat{x}_{ti}}$$

$$\underline{1+x+x^2-e^x} \geq 0 \quad \forall x \leq 1$$

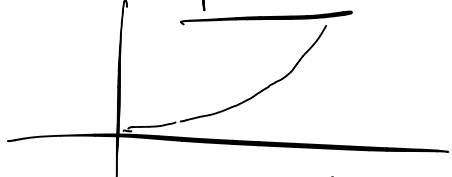
$$1+2x-e^x$$

$$e^{-3} =$$

$$\frac{\partial}{\partial^2} : \quad 2-e^x \qquad e^x$$



$$1+x+x^2$$



$$1-1+1$$
$$= 1$$

$$1+x+x^2$$

Next:

Want regret to be small in expectation but with high probability!

Will show: $\forall \delta \in (0,1)$, w.p $(1-\delta)$,

$$\hat{R}_n = O\left(\sqrt{nk\log\left(\frac{k}{\delta}\right)}\right)$$

Modification: $\hat{Y}_{ti} = \dfrac{\mathbb{1}\{A_t = i\} Y_t}{P_{ti} + \gamma}$

Called EXP-3-IX, implicit exploration