Exp3 in the Linear Setting

Source: Bandit Algorithms, L. & S., Chap 27.

Adversarial Setting for Linear Bandits.

Setting: Actions $\mathcal{A} \subseteq \mathbb{R}^d$, $t = 1, 2, \ldots, n$

Player: Choose $A_t \in \mathcal{A}$, and receives reward

$\in \mathbb{R}^d$

$Y_t = \langle A_t, y_t \rangle$.

loss vector chosen by adversary; $y_t \in \mathbb{R}^d$

Adversary: As before, adversary aware of Player's policy (but not action, as it is randomized using secret bits; see Set 8 – Exp3 Algorithm).

Assumptions:

① $\forall\, t = 1, 2, \ldots, n$, $y_t \in \mathcal{L} = \{ z \in \mathbb{R}^d : \sup_{a \in \mathcal{A}} |\langle a, y \rangle| \leq 1 \}$

② $\mathcal{A}$ spans $\mathbb{R}^d$

Regret: $R_n = E\left[\sum_{t=1}^{n} Y_t\right] - \min_{a \in \mathcal{A}} \sum_{t=1}^{n} \langle a, y_t \rangle$

Linear Exp3 Algorithm Idea: (Supposing $|\mathcal{A}|$ finite)

(a) Based on $Y_t \in \mathbb{R}$, construct $\hat{Y}_t \in \mathbb{R}^d$, which is an unbiased estimate for $y_t \in \mathbb{R}^d$.

$\boxed{\text{(observed loss which is the projection of } y_t \text{ along } A_t\text{)}}$

(b) Use $\hat{Y}_t$ to exponentially reweight prob. distribution over $a \in \mathcal{A}$.

(c) Play (using secret randomness) a randomly chosen arm $a \in \mathcal{A}$ using this distribution.

Assumption: $|\mathcal{A}|$ is finite (for now).

Importance Sampling Estimator:

prob. with which arm a played at time t

$$\hat{Y}_t = Q_t^{-1} A_t Y_t, \quad \text{where } Q_t = \sum_{a \in \mathcal{A}} P_t(a)\, a a^T,$$

$$\hat{Y}_t(a) = \langle a, \hat{Y}_t \rangle.$$

$$P_t(a) = \gamma \pi(a) + (1-\gamma) \frac{e^{-\eta \sum_{s=1}^{t-1} \hat{Y}_s(a)}}{\sum_{b \in \mathcal{A}} e^{-\eta \sum_{s=1}^{t-1} \hat{Y}_s(b)}}$$

an exploration distribution that is used for variance control, π chosen using Kiefer-Wolfowitz Thm

Parameters: We will see $\gamma = g(\pi) \cdot \eta$, $\eta \sim \sqrt{\frac{\log(k)}{(2g(\pi)+d)n}}$

Kiefer-Wolfowitz

$g(\pi) \leq d$

Algorithm: Input $\mathcal{A} \subseteq \mathbb{R}^d$, η the learning rate, exploration dist. π, exploration parameter γ.

For each $t = 1, 2, \ldots, n$:

① Compute $P_t(a)$ as above, and play arm $A_t \sim P_t$.

② Update $\hat{Y}_t = Q_t^{-1} A_t Y_t = Q_t^{-1} A_t \langle A_t, y_t \rangle$

with $\hat{Y}_t(a) = \langle a, \hat{Y}_t \rangle$

Theorem (27.1 in text): Setting as above. Then $\exists\, \gamma, \eta$ s.t. for any π,

$$R_n \leq 2\sqrt{(2g(\pi)+d)\, n \log k},$$

where $k = |\mathcal{A}|$, $g(\pi) = \max_{a \in \mathcal{A}} \|a\|^2_{\underbrace{Q(\pi)^{-1}}}$

$$Q(\pi) = \sum_{a \in \mathcal{A}} \pi(a) a a^T$$

Further, $\exists\, \pi$ s.t. $g(\pi) \leq d$ and

$$R_n \leq 2\sqrt{3 d n \ln(k)}$$

Proof: Similar in spirit to Exp3 proof. Assume η small enough s.t. $-1 \leq \eta \hat{Y}_t(a) \leq 1 \quad \forall a \in \mathcal{A}$.

$\hookrightarrow$ explicit value later in proof.

Using Exp3 proof, we can get (analog of Step 1):

$$R_n \leq \frac{\ln(k)}{\eta} + 2\gamma n + \eta \sum_{t=1}^{n} \underbrace{E\left[\sum_{a \in \mathcal{A}} P_t(a) \hat{Y}_t^2(a)\right]}_{①}$$

Bounding $E\Big[\underbrace{\sum_{a \in \mathcal{A}} P_t(a) \hat{Y}_t^2(a)}_{\triangleq M_t}\Big]$:

$$\hat{Y}_t^2(a) = \left(\langle a, \hat{Y}_t \rangle\right)^2 = \left(\langle a, Q_t^{-1} A_t Y_t \rangle\right)^2$$

$$= \left(a^T Q_t^{-1} A_t \underbrace{Y_t}_{\text{scalar}}\right)^2 = Y_t^2 \left(A_t^T Q_t^{-1} a a^T Q_t^{-1} A_t\right)$$

$$\therefore M_t = \sum_{a \in \mathcal{A}} P_t(a) \hat{Y}_t^2(a)$$

$$= Y_t^2 A_t^T Q_t^{-1} \underbrace{\left(\sum_{a \in \mathcal{A}} P_t(a) a a^T\right)}_{Q_t} Q_t^{-1} A_t$$

$$= \underbrace{Y_t^2}_{\leq 1} A_t^T Q_t^{-1} A_t \leq A_t^T Q_t^{-1} A_t$$

$$= \text{trace}\left(A_t^T Q_t^{-1} A_t\right)$$

$$= \text{trace}\left(A_t A_t^T Q_t^{-1}\right)$$

$$\therefore E\left[M_t \mid A_1, A_2, \cdots A_{t-1}\right] \leq \text{trace}\left(\underbrace{\left(\sum_{a \in \mathcal{A}} P_t(a) a a^T\right)}_{Q_t} Q_t^{-1}\right)$$

$$= d.$$

Recall: $|\eta \hat{Y}_t(a)| \leq 1$, and $|Y_t| \leq 1$. Thus,

$$|\eta \hat{Y}_t(a)| = |\eta a^T Q_t^{-1} A_t Y_t| \leq \eta |a^T Q_t^{-1} A_t| \quad (*)$$

$\boxed{\text{Recall} \quad Q(\pi) = \sum_{a \in \mathcal{A}} \pi(a) a a^T}$

Now, we have:

$$Q_t = \sum_{a \in \mathcal{A}} \underbrace{P_t(a)}_{\gamma \pi(a) + (1-\gamma)(\cdot)} a a^T \succeq \gamma Q(\pi)$$

$$\therefore \quad Q_t^{-1} \preceq \frac{Q(\pi)^{-1}}{\gamma}.$$

Thus, $|a^T Q_t^{-1} A_t| = |\langle a, Q_t^{-1} A_t \rangle| \leq \|a\|_{Q_t^{-1}} \|Q_t^{-1} A_t\|_{Q_t}$

$$= \|a\|_{Q_t^{-1}} \cdot (A_t^T Q_t^{-1} Q_t Q_t^{-1} A_t) = \|a\|_{Q_t^{-1}} \|A_t\|_{Q_t^{-1}}$$

$$\leq \max_{v \in \mathcal{A}} v^T Q_t^{-1} v \leq \frac{1}{\gamma} \max_{v \in \mathcal{A}} v^T Q^{-1}(\pi) v = \frac{g(\pi)}{\gamma}.$$

recall set 16: Kiefer-Wolfowitz Thm

Thus, $|\eta \hat{Y}_t(a)| \leq \frac{\eta}{\gamma} g(\pi).$

Choose $\gamma = \eta g(\pi) \implies |\eta \hat{Y}_t(a)| \leq 1.$

$\therefore$ Substituting in ①:

$$R_n \le \frac{\ln(k)}{\eta} + \underset{\big\downarrow\ \eta g(\pi)}{2\gamma n} + \eta \sum_{t=1}^{n} \underbrace{\mathbb{E}\left[\sum_{a\in A} P_t(a)\, \hat{Y}_t^{2}(a)\right]}_{\le d}$$

$$\le \frac{\ln(k)}{\eta} + \eta n \left(2g(\pi) + d\right)$$

$$\eta = \sqrt{\frac{\ln(k)}{2g(\pi) + d}}$$

$$= 2\sqrt{(2g(\pi)+d)\, n \ln(k)} \qquad \blacksquare$$