CS 7545: Machine Learning Theory **Fall 2018**

Lecture 14: Multi-armed Bandit + EXP3

Lecturer: Jacob Abernethy *Scribes: Ting Gu, Zhanzhan Zhao*

14.1 Multi-armed Bandit Problem

14.1.1 Bandit Setting

Imagine you have a slot machine with multiple arms you can pull. The arms give you reward, but you do not know ahead of time which arm is the best one. You will get feedback from the arm you pulled, but you do not get feedback from arms you did not pull.

- Expert setting: full information feedback!

  You pick one or a combination of expert each round, but at the end you get to see the full loss vector. In hindsight, you can always compute the performance alternatives, which is standard machine learning setting.

- Bandit setting: feedback limited to chosen action.

  Your feedback is limited to the action took. You do not know what would have happened if you took another action, which is more like life.

14.1.2 Protocol

You have n actions/arms
For $t = 1, ..., T$:

- Algorithm selects (usually randomly, but does not have to) $i_t \in [n]$.
- Nature reveals $\ell^t_{i_t}$ from unobserved vector $\ell^t \in [0, 1]^n$.
- Algorithm pays $\ell^t_{i_t}$.
- Algorithm updates accordingly.

14.1.3 Two standard settings:

- Adversarial: ℓ^t chosen arbitrarily, but fixed in advance (discussed in this lecture)
- Stochastic: $\ell^t \overset{IID}{\sim} D$

14.2 The EXP3 Algorithm

The Exp3 Algorithm (2002, Auer, Cesa-Bianchi, Freund, and Schapire) was invented to solve multi-armed bandit problem. It initially proves $\text{Reg}_T = O(T^{\frac{2}{3}})$, later proves $\text{Reg}_T = O(T^{\frac{1}{2}})$. It is basically EWA but with one little trick to handle limited information.

Extension

- Bandit OLO(Online Linear Optimization): AHR '08 efficient $\sqrt{n^2T}$.
- Bandit OCO(Online Convex Optimization): BE '16 efficient $n^{16}\sqrt{T}$.

Key trick unbiased estimation from 1 point

- Let $i_t \sim p^t$, observe $\ell^t_{i_t}$ (but not the rest of the ℓ^t).
- Estimate $\hat{\ell}^t = (0, ...0, \frac{\ell^t_I}{p^t_I}, 0...0)$ ($\frac{\ell^t_I}{p^t_I}$ is at the I-th coordinate if $i_t = I$, 0 everywhere else).

  So

$$\mathop{\mathbb{E}}_{I\sim p_t}[\hat{\ell}_t] = \sum_{i=1}^{n} Pr[i_t = i]\hat{\ell}_t = \sum_{i=1}^{n} p^t_i(0, ...0, \frac{\ell^t_i}{p^t_i}, 0, ...0) = \ell^t$$

Algorithm 1 The EXP3 Algorithm

Set $w^1_i = 1$ for $i = 1, ..., n$
for $t = 1, ...T$ **do**
  $p^t = \frac{w^t}{\|w^t\|_1}$
  Sample $i_t \sim p^t$
  Observe $\ell^t_{i_t}$
  Estimate $\hat{\ell}^t = (0, ...0, \frac{\ell^t_{i_t}}{p^t_{i_t}}, 0...0)$
  Update $w^{t+1}_i = w^t_i \exp(-\eta\hat{\ell}^t_i)$ (for $i = 1, ..., n$, but actually only updates 1 coordinate)
end for

Facts

- Previously we have $e^{sx} \le 1 + (e^s - 1)x \qquad x \in [0, 1]$
- Now we have $e^{-x} \le 1 - x + \frac{x^2}{2} \qquad x \ge 0$

Theorem 14.1 $\mathop{\mathbb{E}}_{Alg\ randomness}[\sum_{t=1}^{T} \ell^t_{i_t} - \min_{i\in[n]} \sum_{t=1}^{T} \ell^t_i] \le \frac{\log n}{\eta} + \frac{\eta}{2}Tn$

Corollary 14.2 *if* $\eta = \sqrt{\frac{2\log n}{Tn}}$, $\mathbb{E}[\mathrm{Reg}_T] \le \sqrt{2Tn\log n}$

It is not surprising that the regret is worse than the expert setting. Because you have to spend more time to inspect each arm.

There is a lower bound $\mathbb{E}[\mathrm{Reg}_T] \ge 2\sqrt{nT}$(Bubeck *et al* '09).

Abernethy *et al* '15 further tightened the bound of EXP3 Algorithm by using Tsallis entropy.

$$\arg\min_p \eta \sum_{s=1}^{t} \ell^s \cdot p + \sum_{i=1}^{N} p_i \log p_i \Rightarrow p^*_i = \exp(-\eta \sum_{s=1}^{t} \ell^s_i)/Z_t$$

By replacing the Shannon entropy with Tsallis entropy to drop the $\sqrt{\log n}$ term.

$$S(\alpha) = \frac{1}{1-\alpha}(1 - \sum_{i=1}^{N} p_i)$$

Proof: Let $\Phi_t = -\frac{1}{\eta}\log(\sum_{t=1}^n w_i^t)$.
Observe that

$$\begin{aligned}
\Phi_{t+1} - \Phi_t &= -\frac{1}{\eta}\log(\frac{\sum_{i=1}^n w_i^{t+1}}{\sum_{i=1}^n w_i^t}) = -\frac{1}{\eta}\log(\frac{\sum_{i=1}^n w_i^t \exp(-\eta\hat{\ell}_i^t)}{\sum_{i=1}^n w_i^t}) \\
&= -\frac{1}{\eta}\log \mathbb{E}_{I\sim p^t}[\exp(-\eta\hat{\ell}_I^t)] \\
&\geq -\frac{1}{\eta}\log \mathbb{E}_{I\sim p^t}[1 - \eta\hat{\ell}_I^t + \frac{1}{2}(\eta\hat{\ell}_I^t)^2] \\
&= -\frac{1}{\eta}\log(1 - \mathbb{E}_{I\sim p^t}[\eta\hat{\ell}_I^t - \frac{1}{2}(\eta\hat{\ell}_I^t)^2]) \\
&\geq \frac{1}{\eta}\mathbb{E}_{I\sim p^t}[\eta\hat{\ell}_I^t - \frac{1}{2}(\eta\hat{\ell}_I^t)^2] \\
&= \sum_{i=1}^n p_i^t\hat{\ell}_i^t - \frac{\eta}{2}\sum_{i=1}^n p_i^t(\hat{\ell}_i^t)^2 \\
&= p^t\hat{\ell}^t - \frac{\eta}{2}\sum_{i=1}^n p_i^t(\hat{\ell}_i^t)^2,
\end{aligned}$$

where $p^t = [p_1^t \dots p_n^t]^{\mathrm{T}}$, and $\hat{\ell}^t = [\hat{\ell}_1^t \dots \hat{\ell}_n^t]$. Recall that $\mathbb{E}_{I\sim p^t}[\hat{\ell}^t] = \ell^t$.
Therefore

$$\begin{aligned}
\mathbb{E}_{i_t\sim p^t}[\Phi_{t+1} - \Phi_t | i_1, \dots, i_{t-1}] &\geq \mathbb{E}_{i_t\sim p^t}[p^t\hat{\ell}^t - \frac{\eta}{2}\sum_{i=1}^n p_i^t(\hat{\ell}_i^t)^2 | i_1, \dots, i_{t-1}] \\
&= p^t\ell^t - \frac{\eta}{2}\mathbb{E}_{i_t\sim p^t}[p_i^t(\frac{\ell_i^t}{p_i^t})^2] \\
&= p^t\ell^t - \frac{\eta}{2}\sum_{i=1}^n p_i^t[p_i^t(\frac{\ell_i^t}{p_i^t})^2] \\
&\geq p^t\ell^t - \frac{\eta n}{2}.
\end{aligned}$$

Hence, (It follows from Law of total Expectation that $\mathbb{E}[g(Y)] = \mathbb{E}[\mathbb{E}(g(Y)|X)]$)

$$\begin{aligned}
\mathbb{E}_{i_t\sim p^t}[\Phi_{t+1} - \Phi_1] &= \mathbb{E}_{i_t\sim p^t}[\sum_{t=1}^T \Phi_{t+1} - \Phi_t] \\
&= \mathbb{E}_{i_t\sim p^t}[\sum_{t=1}^T \mathbb{E}_{i_t\sim p^t}[\Phi_{t+1} - \Phi_t | i_1, \dots, i_{t-1}]] \\
&\geq \underbrace{\mathbb{E}_{i_t\sim p^t}[\sum_{t=1}^T p^t\ell^t]}_{\mathbb{E}-\text{loss of EXP3}} - \frac{\eta n T}{2}.
\end{aligned}$$

$$\because w_i^{T+1} = \exp\left[-\eta \sum_{t=1}^{T} \hat{\ell}_i^t\right]$$

$$\therefore \Phi_{T+1} = -\frac{1}{\eta}\log(\sum_{i=1}^{n} w_i^{T+1}) \leq -\frac{1}{\eta}\log w_i^{T+1}$$

$$\leq -\frac{1}{\eta}\log(\exp\left[-\eta \sum_{t=1}^{T} \hat{\ell}_i^t\right]) = \sum_{t=1}^{T} \hat{\ell}_i^t,$$

$$\therefore \mathbb{E}_{i_t \sim p^t}[\Phi_{t+1} - \Phi_1] \leq \sum_{t=1}^{T} \hat{\ell}_i^t - \Phi_1 \leq \sum_{t=1}^{T} \hat{\ell}_i^t + \frac{\log n}{\eta}$$

Putting all together, it follows that

$$\mathbb{E}_{i_t \sim p^t}[\sum_{t=1}^{T} p^t \ell^t] - \frac{\eta n T}{2} \leq \mathbb{E}_{i_t \sim p^t}[\Phi_{t+1} - \Phi_1] \leq \sum_{t=1}^{T} \hat{\ell}_i^t + \frac{\log n}{\eta},$$

$$\therefore \mathbb{E}_{i_t \sim p^t}[\sum_{t=1}^{T} p^t \ell^t] - \sum_{t=1}^{T} \hat{\ell}_i^t \leq \frac{\log n}{\eta} + \frac{\eta n T}{2},$$

$$\therefore \mathbb{E}[\mathrm{Reg}_T(\mathrm{EXP3})] \leq \frac{\log n}{\eta} + \frac{\eta n T}{2}$$

We choose $\eta = \sqrt{\frac{2\log n}{Tn}}$, then it follows from Corollary 14.2 that

$$\mathbb{E}[\mathrm{Reg}_T(\mathrm{EXP3})] \leq \sqrt{2Tn\log n}$$

.

■