Exp3 and Variants

April 8, 2021

Exp 3

In this lecture we consider the adversarial setting for multi-armed bandits. In each round an adversary chooses a loss vector ℓ_t, the learner picks an index $I_t \in [k]$ and observes $\ell_{t,i}$. The regret of the learner is given by,

$$R_t = \sum_{s=1}^{t} \ell_{s,I_s} - \min_{i\in[k]} \sum_{s=1}^{t} \ell_{si}$$

In general, we can consider two types of adversaries,

- *arbitrary:* where the adversary chooses it's loss vector based on the previous actions of the player
- *oblivious:* where the set of losses is chosen ahead of time.

In general the regret is a random variable so we would like to consider the expectation,

$$\mathbb{E}R_s = \mathbb{E}\left[\sum_{s=1}^{t} \ell_{s,I_s} - \min_{i\in[k]} \sum_{s=1}^{t} \ell_{s,i}\right] = \max_{i\in[k]} \mathbb{E}[\sum_{s=1}^{t} \ell_{s,I_s} - \ell_{s,i}]$$

where the expectation is over the randomness in the learner and the adversary. However it's easier to start with the pseudo-regret,

$$PR_s = \mathbb{E}[\sum_{s=1}^{t} \ell_{s,I_s} - \min_{i\in[k]} \mathbb{E} \sum_{s=1}^{t} \ell_{s,i}]$$

And so, $PR_s \leq \mathbb{E}R_s$
Note that for an oblivious adversary,

$$\min_{i\in[k]} \mathbb{E} \sum_{s=1}^{t} \ell_{i,s} = \min_{i\in[k]} \sum_{s=1}^{t} \ell_{i,s}$$

so the expected regret is the same as the pseudo-regret.
At each round t we only see the reward of the arm I_t that was played. You can show that any deterministic policy will necessarily incur linear regret so moving forward we will consider only randomized policies. In particular, in round t we will choose a probability vector $q_t \in \Delta_k$ and select $I_t \sim p_t$. This random selection helps us be robust against an adversary by letting us "surprise them" but also by allowing us to build unbiased rewards of each arm. In particular, consider the importance weighted estimator,

$$\hat{\ell}_{t,i} = \frac{\mathbf{1}(I_t = i)\ell_{I_t}}{p_{t,i}}$$

then

$$\begin{aligned}
\mathbb{E}[\hat{\ell}_{t,i}] &= \mathbb{E}[\mathbb{E}[\hat{\ell}_{t,i}|I_1, \ell_{1,I_1}, \cdots, I_{t-1}, \ell_{I_{t-1},\ell_{t-1,I_{t-1}}}]] \\
&= \mathbb{E}[p_{t,i} \cdot \frac{\ell_{t,i}}{p_t} + (1-p_{t,i}) \cdot 0] \\
&= \mathbb{E}[\ell_{t,i}]
\end{aligned}$$

In particular, we have the following estimator for the total reward,

$$\hat{S}_{t,i} = \sum_{s=1}^{t} \hat{\ell}_{s,i}$$

Algorithm 1 EXP3

Input: $\gamma \in [0,1], \eta > 0, k, t$

1: Initialize $\lambda = (\frac{1}{k}, \cdots, \frac{1}{k}), p_0 = \lambda$.
2: **for** $s = 1, \cdots, t$ **do**
3: Let $q_s = (1-\gamma)p_s + \gamma\lambda$
4: Draw $I_s \sim q_s$ and observe loss ℓ_{s,I_s}
5: Calculate the estimated total rewards for each $i \in [k]$

$$\hat{S}_{si} = \sum_{r=1}^{s} \frac{\mathbf{1}(I_r = i)\ell_{r,I_r}}{q_{ri}}$$

.

6: Calculate the sampling distribution

$$p_{s+1,i} = \frac{\exp(-\eta\hat{S}_{si})}{\sum_{j=1}^{k}\exp(-\eta\hat{S}_{sj})} \text{ for } i \in [k].$$

7: **end for**

Theorem 1. *Assume $\gamma \in [0,1]$, $\eta > 0$ and for all $i \in [k]$ and $s \leq t$, we have $|\ell_{si}| \leq 1$. Then for all $i \in [k]$*

$$\mathbb{E}\left[\sum_{s=1}^{t} \ell_{s,I_s} - \ell_{s,i}\right] \leq \frac{\log(k)}{\eta} + 2\gamma s + \eta\mathbb{E}\left[\sum_{s=1}^{t}\sum_{j=1}^{k} q_{s,j}\psi(-\eta\hat{\ell}_{s,j})\right]$$

where $\psi(x) = e^x - 1 - x$.

Let's instantiate a special case. Take $\lambda = (1/k, \cdots, 1/k)$, $\gamma = \eta k$ and $\eta = \sqrt{3k\log(k)/t}$. Then,

$$|\eta\frac{\ell_{t,j}}{q_{t,j}}| \leq \eta\frac{k}{\gamma} \leq 1$$

In particular, using the fact that $\phi(x) \leq x^2$ for $x \leq 1$, we have

$$\begin{aligned}\frac{\log(k)}{\eta} + 2\eta kt + \eta\mathbb{E}[\sum_{s=1}^{t}\sum_{j=1}^{k} q_{s,j}\hat{\ell}_{s,j}^2] &= \frac{\log(k)}{\eta} + 2\eta kt + \eta\sum_{s=1}^{t}\sum_{j=1}^{k} r_{s,j}^2 \\ &\leq \frac{\log(k)}{\eta} + 3\eta kT \\ &\leq \sqrt{3kt\log(k)}\end{aligned}$$

Proof. Firstly, note that

$$\mathbb{E}_{s-1}\left[\frac{\mathbf{1}(I_t = i)\ell_{t,i}}{q_t}\right] = q_t\frac{\ell_{ti}}{q_t} + (1-q_t)\cdot 0 = \ell_{ti}$$

thus,

$$\mathbb{E}[\hat{S}_{ti}] = \mathbb{E}\left[\sum_{s=1}^{t}\mathbb{E}_{t-1}\left[\frac{\mathbf{1}(I_t = i)\ell_{ti}}{q_t}\right]\right] = \mathbb{E}[\sum_{s=1}^{t}\ell_{si}].$$

In addition define $\hat{S}_t = \sum_{s=1}^t \sum_{j=1}^k q_{tj}\hat{\ell}_{tj}$ and $W_t = \sum_{j=1}^k \exp(-\eta \hat{S}_{t,j})$. Note that,

$$\exp(-\eta \hat{S}_{ti}) \leq \sum_{j=1}^k \exp(-\eta \hat{S}_{tj}) = W_t = W_0 \times \frac{W_1}{W_0} \cdots \frac{W_t}{W_{t-1}} = k \prod_{s=1}^t \frac{W_s}{W_{s-1}}$$

So let's consider this ratio,

$$\begin{aligned}
\frac{W_s}{W_{s-1}} &= \sum_{j=1}^k \frac{\exp(-\eta \hat{S}_{sj})}{W_{s-1}} \\
&= \sum_{j=1}^k \frac{\exp(-\eta \hat{S}_{s-1,j}) \exp(-\eta \hat{\ell}_{sj})}{W_{s-1}} \\
&= \sum_{j=1}^k p_{sj} \exp(-\eta \hat{\ell}_{sj}) && \text{(Note that this is like a log-CGF)} \\
&= \sum_{j=1}^k p_{sj} (\psi(-\eta \hat{\ell}_{sj}) + 1 - \eta \hat{\ell}_{sj}) && \text{(by definition of } \phi\text{)} \\
&= 1 - \eta \sum_{j=1}^k p_{sj} \hat{\ell}_{sj} + \sum_{j=1}^k p_{sj} \psi(-\eta \hat{\ell}_{sj}) \\
&\leq \exp(-\eta \sum_{j=1}^k p_{sj} \hat{\ell}_{s,j} + \sum_{j=1}^k p_{sj} \psi(-\eta \hat{\ell}_{sj}))
\end{aligned}$$

Multiplying through, we see that,

$$\exp(-\eta \hat{S}_{ti}) \leq k \exp\left(-\eta \sum_{s=1}^t \sum_{j=1}^k p_{sj} \hat{\ell}_{s,j} + \sum_{s=1}^t \sum_{j=1}^k p_{sj} \psi(-\eta \hat{\ell}_{s,j})\right)$$

which implies that,

$$\begin{aligned}
\sum_{s=1}^t \sum_{j=1}^k p_{sj} \hat{\ell}_{s,j} &\leq \hat{S}_{ti} + \frac{\log(k)}{\eta} + \frac{1}{\eta} \sum_{s=1}^t \sum_{j=1}^k p_{sj} \psi(-\eta \hat{\ell}_{s,j}) \\
&\leq \hat{S}_{ti} + \frac{\log(k)}{\eta} + \frac{1}{\eta(1-\gamma)} \sum_{s=1}^t \sum_{j=1}^k q_{sj} \psi(-\eta \hat{\ell}_{s,j})
\end{aligned}$$

where the second line followed from noting that $p_{sj} = \frac{q_{sj} - \gamma \lambda_j}{1-\gamma} \leq \frac{q_{sj}}{1-\gamma}$. Ok, we now multiply both sides by $1-\gamma$ and add $\gamma \sum_{s=1}^t \sum_{j=1}^k \lambda_j \hat{\ell}_{t,j}$, we see

$$\sum_{s=1}^t \sum_{j=1}^k q_{sj} \hat{\ell}_{s,j} - \sum_{s=1}^t \hat{\ell}_{s,j} \leq -\gamma \hat{S}_{ti} + \gamma \sum_{s=1}^t \sum_{j=1}^k \lambda_j \hat{\ell}_{t,j} + \frac{\log(k)}{\eta} + \frac{1}{\eta} \sum_{s=1}^t \sum_{j=1}^k p_{sj} \psi(-\eta \hat{\ell}_{s,j})$$

Now note that,

$$\begin{aligned}
\mathbb{E}[-\gamma \hat{S}_{ti} + \gamma \sum_{s=1}^t \sum_{j=1}^k \lambda_j \hat{\ell}_{t,j}] &= \gamma \mathbb{E}[\sum_{s=1}^t \sum_{j=1}^k \lambda_j \hat{\ell}_{t,j} - \hat{\ell}_{s,j}] \\
&\leq \gamma \mathbb{E}[\sum_{s=1}^t \sum_{j=1}^k \lambda_j \ell_{sj} - \ell_{sj}] \\
&\leq 2\gamma T && \text{(Assuming } |\ell_{sj}| \leq 1\text{)}
\end{aligned}$$

at which point the result follows. ∎

Remarks.

- In the full information setting the regret scales like $O(\sqrt{T\log(k)})$. We will prove this when we discuss mirror descent.
- Instead of the exponential function, we could have used a different convex function. This leads to some of the algorithms we will see later in the class.
- We can consider an algorithm with $\gamma = 0$ if $\ell_{it} \in [0,1]$ since $-\eta/p_{it} \leq 0$. In this case we can get a slightly tighter constant since $\psi(x) \leq \frac{1}{2}x^2$ when $x \leq 0$.
- Can get the same regret guarantee as long as $\hat{\ell}_{s,i} \leq 1/p_{s,i}$ for all $s \leq t, i \in [k]$.

EXP3-IX

Algorithm 2 EXP3-IX

Input: $\gamma \in [0,1], \eta > 0, k, t$

Initialize $p_0 = (\frac{1}{k}, \cdots, \frac{1}{k})$.

for $s = 1, \cdots, t$ **do**

Draw $I_s \sim p_s$ and observe loss ℓ_{s,I_s}

Calculate the estimated total rewards for each $i \in [k]$

$$\hat{S}_{si} = \sum_{r=1}^{s} \frac{\mathbf{1}(I_r = i)\ell_{r,I_r}}{p_{ri} + \gamma}$$

Calculate the sampling distribution

$$p_{s+1,i} = \frac{\exp(-\eta \hat{S}_{si})}{\sum_{j=1}^{k} \exp(-\eta \hat{S}_{sj})} \text{ for } i \in [k].$$

end for

Theorem 2. *Assuming that* $\ell_t \in [0,1]^k$, $\eta = \sqrt{\frac{2\log(k)}{kt}}$, $\gamma = \sqrt{\frac{3\log(2k/\delta)}{kt}}$ *and* $t \geq klog(k)/2$ *then with probability greater than* $1-\delta$

$$\max_{i\in[k]} \sum_{s=1}^{t} \ell_{s,I_s} - \ell_{s,i} \leq O(\sqrt{kt\log(/\delta)})$$

Proof. To make things a bit easier to parse, we vectorize all sums over k e.g. $\ell_t = (\ell_{t,1}, \cdots, \ell_{t,k})$.
Then

$$\begin{aligned}
\sum_{s=1}^{t} \ell_{s,I_s} - \ell_{s,i} &= \sum_{s=1}^{t} (e_{I_s} - e_i)^\top \ell_s \\
&= \sum_{s=1}^{t} (e_{I_s} - p_s)^\top \ell_s + (p_s - e_i)^\top \ell_s \\
&= \sum_{s=1}^{t} (e_{I_s} - p_s)^\top \ell_s + (p_s - e_i)^\top \hat{\ell}_s + p_s^\top (\ell_s - \hat{\ell}_s) + e_i^\top (\hat{\ell}_s - \ell_s) \\
&\leq \sum_{s=1}^{t} (e_{I_s} - p_s)^\top \ell_s + \sup_{i\in[n]} \sum_{s=1}^{t} (p_s - e_i)^\top \hat{\ell}_s + \sum_{s=1}^{t} p_s^\top (\ell_s - \hat{\ell}_s) + \sup_{i\in[n]} \sum_{s=1}^{n} (\hat{\ell}_{s,i} - \ell_{s,i})
\end{aligned}$$

We will analyze each term separately. Term 2 should come out of the guarantees of Exp3 (see remark above), whereas Terms 1,3,4 can be approached using concentration since our new estimator guarantees that $\hat{\ell}_{i,t}$ is bounded by $1/\gamma$.
Term 2:

$$\begin{aligned}
\sum_{s=1}^{t}(p_s - e_i)^\top \hat{\ell}_s &= \sum_{s=1}^{t}\left(\sum_{j=1}^{k} p_{s,j}\hat{\ell}_{s,i}\right) - \hat{\ell}_{s,i} \\
&\leq \frac{\log(k)}{\eta} + \frac{\eta}{2}\sum_{s=1}^{t}\sum_{j=1}^{k} p_{sj}\hat{\ell}_{s,j}^2 \\
&\leq \frac{\log(k)}{\eta} + \frac{\eta}{2}\sum_{s=1}^{t}\sum_{j=1}^{k} \hat{\ell}_{sj} \\
&\leq \frac{\log(k)}{\eta} + \frac{\eta k t}{2} + \frac{\eta}{2}\sum_{s=1}^{t}\sum_{j=1}^{k} (\hat{\ell}_{sj} - \ell_{sj}) \\
&\leq \frac{\log(k)}{\eta} + \frac{\eta k t}{2} + \frac{k\eta}{2}\max_{j\in[k]}\sum_{s=1}^{t} \hat{\ell}_{sj} - \ell_{sj}
\end{aligned}$$

To see the second to third line inequality note

$$p_{s,i}\hat{\ell}\ell_{s,i}^2 \leq \frac{p_{s,i}\mathbf{1}\{I_s = i\}\ell_{s,i}^2}{(p_{s,i}+\gamma)^2} \leq \frac{\mathbf{1}\{I_s = i\}\ell_{s,i}}{p_{s,i}+\gamma}$$

Finally, if $t \geq k\eta/2$ with the choice of $\eta = \sqrt{\frac{2\log(k)}{kt}}$ we see this is bounded by

$$\sqrt{2kt\log(k)} + \text{Term 4}$$

Term 1:

Lemma 1 (Azuma-Hoeffding Lemma)**.** *Fix any $\lambda > 0$, and $\delta \in (0,1)$. Let X_t be a random process wrt a filtration $\mathcal{F}_t$ with $\mu_t = \mathbb{E}[X_t|F_{t-1}]$ and assume $\mathbb{E}[\exp(\lambda(X_t - \mu_t))|F_{t-1}] \leq \exp(R\lambda^2/2)$. Then, $\sum_{t=1}^{T}(X_t - \mu_t) \leq \sqrt{2RT\log(1/\delta)}$ with probability at least $1-\delta$.*

Let $F_s = (I_1, \ell_{1,I_1}, \cdots, I_s, \ell_{s,I_s})$ then

$$\mathbb{E}[(e_{I_s} - p_s)^\top \ell_s | F_{s-1}] = \mathbb{E}\left[\ell_{s,I_s} - \sum_{j=1}^{k} p_{s,j}\ell_{s,j} | F_{s-1}\right] = 0$$

and $\ell_{s,I_s} \in [0,1]$ so ℓ_{s,I_s} is (conditionally) 1/4-subGaussian so

$$\sum_{s=1}^{t}(e_{I_s} - p_s)^\top \ell_t \leq \sqrt{T\log(4/\delta)/2}$$

with probability greater than $1-\delta/4$
Term 3:

Lemma 2 (Friedman's Inequality)**.** *Fix any $\lambda > 0$, and $\delta \in (0,1)$. Let X_t be a random process wr.t a filtration $\mathcal{F}_t$ with $\mu_t = \mathbb{E}[X_t|F_{t-1}]$ and $V_t = \mathbb{E}[X_t^2|F_{t-1}]$ and assume $\lambda X_t \leq 1$. Then with probability at least $1-\delta$, we have for all t,*

$$\sum_{s=1}^{t} X_s - \mu_s \leq \lambda\sum_{s=1}^{t} V_s + \frac{log(1/\delta)}{\lambda}$$

$$
\begin{aligned}
\sum_{s=1}^{t}\sum_{j=1}^{k} p_{sj}\ell_{sj} - p_{sj}\hat{\ell}_{sj} &= \sum_{s=1}^{t}\sum_{j=1}^{k} p_{sj}(\ell_{sj} - \mathbb{E}[\hat{\ell}_{sj}|F_{t-1}]) + \sum_{s=1}^{t}\sum_{j=1}^{k} p_{sj}(\mathbb{E}[\hat{\ell}_{s,j}|F_{t-1}] - \hat{\ell}_{sj}) \\
&= \sum_{s=1}^{t}\sum_{j=1}^{k} p_{sj}(\ell_{sj} - \frac{p_{sj}\ell_{sj}}{p_{sj}+\gamma}) + \sum_{s=1}^{t}\sum_{j=1}^{k} p_{sj}(\mathbb{E}[\hat{\ell}_{s,j}|F_{t-1}] - \hat{\ell}_{sj}) \\
&= \sum_{s=1}^{t}\sum_{j=1}^{k} p_{sj}\frac{\gamma\ell_{sj}}{p_{sj}+\gamma} + \sum_{s=1}^{t} p_s^\top(\mathbb{E}[\hat{\ell}_s|F_{t-1}] - \hat{\ell}_s) \\
&\leq \gamma kt + \sum_{s=1}^{t} p_s^\top(\mathbb{E}[\hat{\ell}_s|F_{t-1}] - \hat{\ell}_s)
\end{aligned}
$$

Now, note that $p_t^\top(\mathbb{E}[\hat{\ell}_s|F_{t-1}] - \hat{\ell}_s) \leq 1$ and

$$\mathbb{E}[(p_t^\top\hat{\ell}_s)^2] \leq 1$$

so applying Friedman's inequality with probability greater than $1-\delta/4$

$$\sum_{s=1}^{t}\sum_{j=1}^{k} p_{sj}\ell_{sj} - p_{sj}\hat{\ell}_{sj} \leq \gamma kt + \gamma t + \gamma^{-1}\log(4/\delta)$$

Term 4 This is the interesting one to bound and all proofs of high probability bounds depend on it to some degree. By the computation in the second equality above,

$$
\begin{aligned}
\mathbb{E}[\hat{\ell}_{si}|F_{s-1}] - \ell_{si} &= \frac{-\gamma\ell_{si}}{p_{si}+\gamma} \\
\mathbb{E}[\hat{\ell}_{si}^2|F_{s-1}] &= p_{si}\frac{\ell_{si}^2}{(p_{si}+\gamma)^2} \leq \frac{\ell_{si}}{p_{si}+\gamma}
\end{aligned}
$$

Thus we can apply Freedman's inequality to the first term

$$
\begin{aligned}
\sum_{s=1}^{t}(\hat{\ell}_{si} - \mathbb{E}[\hat{\ell}_{si}|F_{s-1}] + \mathbb{E}[\hat{\ell}_{si}|F_{s-1}] - \ell_{si}) &= \sum_{s=1}^{t}(\hat{\ell}_{si} - \mathbb{E}[\hat{\ell}_{si}|F_{s-1}]) - \sum_{s=1}^{t}\frac{\gamma\ell_{si}}{p_{si}+\gamma} \\
&\leq \left[\gamma\left(\sum_{s=1}^{t}\frac{\ell_{si}}{p_{si}+\gamma}\right) + \frac{\log(2k/\delta)}{\gamma}\right] - \sum_{s=1}^{t}\frac{\gamma\ell_{si}}{p_{si}+\gamma} \\
&= \frac{\log(2k/\delta)}{\gamma}
\end{aligned}
$$

Where in the application of Bernstein we have considered a union bound over the k arms. In EXP3, this quantity is not so well controlled! Let's finally combine things,

$$\sum_{s=1}^{t}\ell_{t,I_t} - \ell_{t,i} \leq \sqrt{T\log(4/\delta)/2} + \sqrt{2kT\log(k)} + \gamma(k+1)t + \gamma^{-1}\log(4/\delta) + 2\gamma^{-1}\log(2k/\delta)$$

choosing $\gamma = \sqrt{\frac{3\log(2n/\delta)}{nt}}$