ExpByOpt for Convex Bandits

June 3, 2025

1 Setup and Notation

Reuse the notation from Lattimore [2024].

Assumption 1. The following hold:

- The losses are in $\mathcal{F}_b$.
- There is no noise so that $Y_t = f_t(X_t)$.

Assumption 2. $\mathcal{C}$ is finite subset of K such that:

- $\log(\mathcal{C}) \leq \tilde{\mathcal{O}}(d)$.
- For all $f \in \mathcal{F}_b$ there exists $x \in \mathcal{C}$ such that $f(x) \leq \inf_{x' \in K} f(x') + \frac{1}{n}$.

2 Exponential Weights with Importance Sampling

Let $(\hat{s}_t)_{t=1}^n : \mathcal{C} \to \mathbb{R}$ be a sequence of functions and

$$q_t(x) = \frac{\exp(-\eta \sum_{u=1}^{t-1} \hat{s}_u(x))}{\sum_{y \in \mathcal{C}} \exp(-\eta \sum_{u=1}^{t-1} \hat{s}_u(y))}, \quad x \in \mathcal{C}\,. \tag{1}$$

The following theorem gives a bound on the regret of the exponential weights algorithm in the original setting. *Here, it's better to think of $\hat{s}_t$ as the losses observed at time t and not the estimates of the losses.*

Theorem 3 (Lattimore [2024], Theorem 8.11)**.** *For any $y \in \mathcal{C}$ we have*

$$\sum_{t=1}^n \langle q_t, \hat{s}_t \rangle - \hat{s}_t(y) \leq \frac{\log(|\mathcal{C}|)}{\eta} + \frac{1}{\eta} \sum_{t=1}^n \mathcal{S}_t(\eta \hat{s}_t)\,,$$

where $\mathcal{S}_t(u) = D_{R^\star}(R'(q_t) - u, R'(q_t))$.

The following theorem applies to the setting when $\langle f_t, X_t \rangle$ is observed and not f_t. The idea is to use importance sampling and estimates $\hat{s}_t$ to compute the losses.

```
args: learning rate η > 0
let C ∈ K be finite
for t = 1 to n:
    compute q_t(x) = exp(-η Σ_{u=1}^{t-1} ŝ_u(x)) / Σ_{y∈C} exp(-η Σ_{u=1}^{t-1} ŝ_u(x)) for all x ∈ C
    find distribution p_t as a function of q_t
    sample X_t ~ p_t , and observe Y_t = f_t(X_t)
    compute ŝ_t(x) ∀x ∈ C using p_t, q_t, X_t, Y_t
```

Algorithm 1: Exponential Weights with Importance Sampling

Theorem 4 (Lattimore [2024], Theorem 8.14)**.** *Let $x_\star = \arg\min_{x\in\mathcal{C}} \sum_{t=1}^{n} f_t(x)$ and $p_\star \in \Delta(\mathcal{C})$ be a Dirac on $x_\star$. The expected regret of Algorithm 1 is bounded by*

$$\mathbb{E}\left[\mathfrak{R}_n(x_\star)\right] = \frac{\log(|\mathcal{C}|)}{\eta} + \sum_{t=1}^{n} \mathbb{E}\left[\langle p_t - p_\star, f_t\rangle + \langle p_\star - q_t, \hat{s}_t\rangle + \frac{1}{\eta}\mathcal{S}_t(\eta\hat{s}_t)\right].$$

Proof. Immediate from Theorem 3. □

3 Exploration By Optimisation

Let

- $\mathcal{G}$ be the set of all functions $g : \mathcal{C} \to \mathbb{R}$, i.e., $\mathcal{G} = \mathbb{R}^{|\mathcal{C}|}$.
- $\mathcal{E}$ be the set of all functions $E : \mathcal{C} \times \mathbb{R} \to \mathcal{G}$.

The idea is to bound the term inside the expectation in Theorem 4 uniformly over all $t \in [n]$. To this end, define

$$\Lambda_\eta(q, p, E, r, f) = \frac{1}{\eta}\mathbb{E}\left[\langle p - r, f\rangle + \langle r - q, E(X, Y)\rangle + \frac{1}{\eta}\mathcal{S}_q(\eta E(X, Y))\right].$$

Note that the randomness is only through X and Y, where $X \sim p$ and $Y = f(X)$. For the learner, the degree of freedom is in the choice of the estimator $\hat{s}_t$ and the exploration distribution p_t. Therefore, fix $t \in [n]$ and define

$$\Lambda_\eta(q) = \inf_{p\in\Delta(\mathcal{C}), E\in\mathcal{E}} \ \sup_{r\in\Delta(\mathcal{C}), f\in\mathcal{F}_b} \frac{1}{\eta}\Lambda_\eta(q, p, E, r, f).$$

Remark 5. There is a bit of waste here since the supremum is taken over all $r \in \Delta(\mathcal{C})$, but it could've been restricted to dirac distributions.

It's worth noting that Algorithm 2 is a instantiation of Algorithm 1 where the estimator $\hat{s}_t$ and the exploration distribution p_t are chosen by solving an optimisation problem.

```
args: learning rate η > 0, precision ϵ > 0
let C ∈ K be finite
for t = 1 to n:
    compute q_t(x) = exp(−η ∑_{u=1}^{t−1} ŝ_u(x)) / ∑_{y∈C} exp(−η ∑_{u=1}^{t−1} ŝ_u(x)) for all x ∈ C
    find distribution p_t ∈ Δ(C) and E_t ∈ ℰ such that
        Λ_η(q_t, p_t, E_t) ≤ inf_{p,E} Λ_η(q_t) + ϵ
    sample X_t ∼ p_t, and observe Y_t = f_t(X_t)
    compute ŝ_t(x) = E_t(X_t, Y_t)
```

Algorithm 2: Exploration by Optimisation

Theorem 6 (Lattimore [2024], Theorem 8.15). *The expected regret of Algorithm 2 is bounded by*

$$\mathbb{E}\left[\mathfrak{R}_n(x_\star)\right] \le \frac{\log(|\mathcal{C}|)}{\eta} + n\eta \sup_{q\in\Delta(\mathcal{C})} \Lambda_\eta(q) + n\eta\epsilon\,.$$

Therefore, from Theorem 6, to bound the regret of Algorithm 2 it suffices to bound

$$\Lambda^\star_{\mathcal{C}} = \sup_{\eta>0, q\in\Delta(\mathcal{C})} \inf_{p\in\Delta(\mathcal{C}), E\in\mathcal{E}} \sup_{y\in\mathcal{C}, f\in\mathcal{F}_b} \Lambda_\eta(q,p,E,y,f)\,.$$

The following lemma shows that the order of the infimum and supremum can be interchanged.

Lemma 7. *For all $\eta > 0$ and $q \in \Delta(\mathcal{C})$, we have*

$$\inf_{p\in\Delta(\mathcal{C}), E\in\mathcal{E}} \sup_{r\in\Delta(\mathcal{C}), f\in\mathcal{F}_b} \Lambda_\eta(q,p,E,r,f) = \sup_{r\in\Delta(\mathcal{C}), f\in\mathcal{F}_b} \inf_{p\in\Delta(\mathcal{C}), E\in\mathcal{E}} \Lambda_\eta(q,p,E,r,f)\,. \tag{2}$$

3.1 Unrestricted estimator class $\mathcal{E}$

The minimizer E in the general case has the following close form.

Lemma 8. *Given $r \in \Delta(\mathcal{C})$, $f \in \mathcal{F}_b$, and $p \in \Delta(\mathcal{C})$ define $G_{r,f,p} \in \mathcal{E}$ as*

$$G_{r,f,p}(x,y) = \frac{1}{\eta}\left(R'(q) - R'(\mathbb{E}[p_\star | f(x) = y])\right)\,, \tag{3}$$

then we have

$$\inf_{E\in\mathcal{E}} \Lambda_\eta(q,p,E,r,f) = \Lambda_\eta(q,p,G_{r,f,p},r,f).$$

3.2 Restricted estimator class $\mathcal{E}$

Define E through a probability kernel T such that

$$E(x,y) = \frac{T(x|\cdot)y}{p(x)} \; (= \hat{s}_x(\cdot))\,,$$
$$\text{and} \quad \sum_{x\in\mathcal{C}} T(x|y) = 1 \quad \text{for all } y \in \mathcal{C}\,,$$

Again, fix $r \in \Delta(\mathcal{C})$, $f \in \mathcal{F}_b$, and $p \in \Delta(\mathcal{C})$. We are interested in the T that minimizes

$$\begin{aligned}\Lambda(T) :=& \mathbb{E}\left[\langle p-r, f\rangle + \langle r-q, E(X,Y)\rangle + \frac{1}{\eta}\mathcal{S}_q(\eta E(X,Y))\right]\\ =& \;\mathbb{E}\left[\mathbb{E}\left[\langle p-r, f\;\rangle + \langle r-q, \tfrac{T(X|\cdot)Y}{p(X)}\rangle + \frac{1}{\eta}\mathcal{S}_q\left(\eta\tfrac{T(X|\cdot)Y}{p(X)}\right)\middle|X\right]\right]\,.\end{aligned}$$

Define the shorthand $T_x = T(X|\cdot) \in \mathbb{R}^{|\mathcal{C}|}$. Then through differentiation we have

$$\begin{aligned}&\nabla_{T_x}\Lambda(T)\\ &= \nabla_{T_x}\mathbb{E}\left[\langle p-r, f\;\rangle + \langle r-q, \tfrac{T(X|\cdot)Y}{p(X)}\rangle + \frac{1}{\eta}\mathcal{S}_q\left(\eta\tfrac{T(X|\cdot)Y}{p(X)}\right)\middle|X=x\right]\\ &= \nabla_{T_x}\mathbb{E}\left[\langle r-q, \tfrac{T_xY}{p(X)}\rangle + \frac{1}{\eta}\left(R^\star\left(R'(q) - \eta\tfrac{T_xY}{p(X)}\right) - R^\star(R'(q)) - \nabla R^\star(R'(q))^\top(-\eta\tfrac{T_xY}{p(X)})\right)\middle|X=x\right]\\ &= \nabla_{T_x}\mathbb{E}\left[\langle r, \tfrac{T_xY}{p(X)}\rangle - \tfrac{Y}{p(X)}q^\top T_x + \frac{1}{\eta}\left(R^\star\left(R'(q) - \eta\tfrac{T_xY}{p(X)}\right) + \tfrac{\eta Y}{p(X)}q^\top T_x\right)\middle|X=x\right]\\ &= \nabla_{T_x}\mathbb{E}\left[\langle r, \tfrac{T_xY}{p(X)}\rangle + \frac{1}{\eta}R^\star\left(R'(q) - \eta\tfrac{T_xY}{p(X)}\right)\middle|X=x\right]\\ &= \mathbb{E}\left[\frac{Y}{p(x)}\Big(r - \nabla R^\star\Big(R'(q) - \eta\,\tfrac{T_x\,Y}{p(x)}\Big)\Big)\,\middle|\,X=x\right]\\ &= \frac{1}{p(x)}\mathbb{E}\left[Y\Big(r - \exp\Big(R'(q) - \eta\,\tfrac{T_x\,Y}{p(x)}\Big)\Big)\,\middle|\,X=x\right]\\ &= \frac{1}{p(x)}\left(f(x)r - \mathbb{E}\left[Y\exp\Big(R'(q) - \tfrac{\eta}{p(x)}YT_x\Big)\,\middle|\,X=x\right]\right)\end{aligned}$$

we want to solve for T_x such that the gradient is zero. Therefore, for coordinate $z \in \mathcal{C}$ we have

$$\mathbb{E}\left[Y\exp\Big(R'(q)_z - \tfrac{\eta}{p(x)}YT_{x,z}\Big)\,\middle|\,X=x\right] = f(x)r_z\,,$$

which is equivalent to

$$\begin{aligned}
&\sum_{x' \in \mathcal{C}} p(x') f(x') \exp\Big(R'(q)_z - \tfrac{\eta}{p(x)} f(x') T_{x,z} \Big) = f(x) r_z \\
&\quad \Leftrightarrow \sum_{x' \in \mathcal{C}} p(x') f(x') \exp\Big(- \tfrac{\eta}{p(x)} f(x') T_{x,z} \Big) = \exp(-R'(q)_z) f(x) r_z
\end{aligned}$$

References

Tor Lattimore. Bandit convex optimisation. *arXiv preprint arXiv:2402.06535*, 2024.