Adaptive Rates for Interactive Decision Making

Atul Ganju, Karthik Sridharan

December 2, 2024

1 Generalizing the Averaged DEC

1.1 Reverse-Engineering the Small Loss Bound

Suppose there is a parameter space Θ which parameterizes a family of distributions $p : \Theta \times \mathcal{X} \times \mathcal{A} \to \Delta(\mathbb{R})$. Define the function $\ell(\theta, x, a) = \mathbb{E}_{L \sim p(\theta,x,a)}[L|x, a]$ to be the mean reward received under parameter θ in response to taking action a after seeing context x. We will then note that the realizability assumption amounts to there being a $\theta_0 \in \Theta$ such that $\ell(\theta_0, X_t, A_t) = \mathbb{E}_{L_t \sim p(\theta,x,a)}[L_t|X_t, A_t]$ for all $t \in [T]$. Then, their bound on regret looks like:

$$
\begin{aligned}
&\mathbb{E}[\mathbf{Reg}(\theta_0)] \\
&= \mathbb{E}\left[\sum_{t=1}^{T} \mathbb{E}_{A_t\sim\pi_t}[\ell_t(\theta_0, A_t)|X_t, \mathcal{H}_{t-1}] - \ell_t^*(\theta_0)\right] \\
&= \mathbb{E}\left[\sum_{t=1}^{T} \overline{\mathrm{DEC}}_{\mu,t}(\pi_t) + \mu \cdot \overline{\mathrm{IG}}_t(\pi_t) + \mathrm{UE}_t + \mathrm{OG}_t\right] \\
&\leq \mathbb{E}\left[\sum_{t=1}^{T} \overline{\mathrm{DEC}}_{\mu,t}(\pi_t) + 4\mu \cdot \mathrm{IG}_t(\pi_t) + \mathrm{UE}_t + \mathrm{OG}_t\right] \\
&\leq \mathbb{E}\left[\sum_{t=1}^{T} \overline{\mathrm{DEC}}_{\mu,t}(\pi_t) + 4\mu \cdot \mathrm{IG}_t(\pi_t) + \left(2\gamma \cdot \mathbb{E}_{A_t\sim\pi_t}[\ell(\theta_0, A_t)|X_t, \mathcal{H}_{t-1}] + \frac{1}{\gamma} \cdot \mathrm{IG}_t(\pi_t)\right) + \mathrm{OG}_t\right] \\
&= \mathbb{E}\left[\sum_{t=1}^{T} \overline{\mathrm{DEC}}_{\mu,t}(\pi_t) + \left(4\mu + \frac{1}{\gamma}\right) \cdot \mathrm{IG}_t(\pi_t) + \mathrm{OG}_t + 2\gamma \cdot \mathbb{E}_{A_t\sim\pi_t}[\ell(\theta_0, A_t)|X_t, \mathcal{H}_{t-1}]\right] \\
&\leq \mathbb{E}\left[\sum_{t=1}^{T} \frac{5}{4} \cdot \left(\overline{\mathrm{DEC}}_{\mu,t}(\pi_t) + \left(4\mu + \frac{1}{\gamma}\right) \cdot \mathrm{IG}_t(\pi_t) + \left(1 - \frac{\lambda\beta}{2}\right) \cdot \mathrm{OG}_t + 2\gamma \cdot \mathbb{E}_{A_t\sim\pi_t}[\ell(\theta_0, A_t)|X_t, \mathcal{H}_{t-1}]\right)\right] \\
&= \mathbb{E}\left[\sum_{t=1}^{T} \frac{5}{4} \cdot \left(\overline{\mathrm{DEC}}_{1/10\lambda,t}(\pi_t) + \frac{1}{2\lambda} \cdot \mathrm{IG}_t(\pi_t) + (1-\lambda) \cdot \mathrm{OG}_t + 20\lambda \cdot \mathbb{E}_{A_t\sim\pi_t}[\ell(\theta_0, A_t)|X_t, \mathcal{H}_{t-1}]\right)\right] \\
&= \mathbb{E}\left[\sum_{t=1}^{T} \frac{5}{4} \cdot \left(\overline{\mathrm{DEC}}_{1/10\lambda,t}(\pi_t) + \frac{1}{2\lambda} \cdot \mathrm{IG}_t(\pi_t) + (1-\lambda) \cdot \mathrm{OG}_t\right) + 25\lambda \cdot \mathbb{E}_{A_t\sim\pi_t}[\ell(\theta_0, A_t)|X_t, \mathcal{H}_{t-1}]\right]
\end{aligned}
$$

which implies,

$$
\mathbb{E}\left[\sum_{t=1}^{T} (1 - 25\lambda) \cdot \mathbb{E}_{A_t\sim\pi_t}[\ell_t(\theta_0, A_t)|X_t, \mathcal{H}_{t-1}] - \ell_t^*(\theta_0)\right]
$$

$$\leq \mathbb{E}\left[\sum_{t=1}^{T} \frac{5}{4} \cdot \left(\overline{\mathrm{DEC}}_{1/10\lambda,t}(\pi_t) + \frac{1}{2\lambda} \cdot \mathrm{IG}_t(\pi_t) + (1-\lambda) \cdot \mathrm{OG}_t\right)\right]$$

Assuming we use the inverse gap weighting strategy, we focus in on the latter two terms in the summation.

$$\begin{aligned}
&\mathbb{E}\left[\sum_{t=1}^{T} \frac{1}{2\lambda} \cdot \mathrm{IG}_t(\pi_t) + (1-\lambda) \cdot \mathrm{OG}_t\right] \\
&= \mathbb{E}\left[\sum_{t=1}^{T} \mathbb{E}_{A_t \sim \pi_t}\left[\mathbb{E}_{\theta \sim Q_t}\left[\frac{1}{2\lambda} \cdot \mathcal{D}^2_{\mathrm{H}}(p_t(L_t|\theta, A_t), p_t(L_t|\theta_0, A_t)) + (1-\lambda) \cdot (\ell^*_t(\theta) - \ell^*_t(\theta_0))\right]\right]\right] \\
&= \mathbb{E}\left[\sum_{t=1}^{T} \mathbb{E}_{A_t \sim \pi_t}\left[-\frac{1}{2\lambda} \cdot \log\left(\mathbb{E}_{\theta \sim Q_t}\left[\left(\frac{p_t(L_t|\theta, A_t)}{p_t(L_t|\theta_0, A_t)}\right)^{1/2}\right]\right) + (1-\lambda) \cdot \mathbb{E}_{\theta \sim Q_t}\left[(\ell^*_t(\theta) - \ell^*_t(\theta_0))\right]\right]\right] \\
&= \mathbb{E}\Bigg[\sum_{t=1}^{T} \mathbb{E}_{A_t \sim \pi_t}\Bigg[-\frac{1}{2\lambda} \cdot \left(\log\left(\mathbb{E}_{\theta \sim Q_t}\left[(p_t(L_t|\theta, A_t))^{1/2}\right]\right) - \log\left((p_t(L_t|\theta_0, A_t))^{1/2}\right)\right) \\
&\quad + (1-\lambda) \cdot \mathbb{E}_{\theta \sim Q_t}\left[(\ell^*_t(\theta) - \ell^*_t(\theta_0))\right]\Bigg]\Bigg].
\end{aligned}$$

To bound this quantity is equivalent to playing an online game with a finite number of experts parameterized by θ. Specifically, for the measurable space $([0,1] \times \mathcal{X} \times \mathcal{A}, \mathcal{B}([0,1] \times \mathcal{X} \times \mathcal{A}))$, take $\mathcal{P}$ to be the set of probability measures over this space and $\mathcal{P}(X) = \{P(\cdot|X = x)|P \in \mathcal{P}, x \in \mathcal{X}\}$. Then the action space of the game is the set of tuples $\{(\sqrt{p}, r) : p \in \mathcal{P}(X), r \in [0,1]\}$ and each expert predicts a tuple of information $((p_t(\cdot, \theta, A_t)^{1/2}, \ell^*_t(\theta))$ parameterized by $\theta \in \Theta$. Finally, the cost function of an action in this space is $c_t((q, r)) = -\log(q) + r$.

Suppose there is a parameter space Θ which parameterizes a family of distributions $p : \Theta \times \mathcal{X} \times \mathcal{A} \to \Delta(\mathbb{R})$. Define the function $\ell(\theta, x, a) = \mathbb{E}_{L \sim p(\theta,x,a)}[L|x, a]$ to be the mean reward received under parameter θ in response to taking action a after seeing context x. We will then note that the realizability assumption amounts to there being a $\theta_0 \in \Theta$ such that $\ell(\theta_0, X_t, A_t) = \mathbb{E}_{L_t \sim p(\theta,x,a)}[L_t|X_t, A_t]$ for all $t \in [T]$. Then one can rewrite regret the small loss bound from the Optimistic Information Directed Sampling paper in the following way,

$$
\begin{aligned}
&\mathbb{E}[\mathbf{Reg}(\theta_0)] \\
&= \mathbb{E}\left[\sum_{t=1}^{T} \mathbb{E}_{A_t \sim \pi_t}[\mathbb{E}_{L_t \sim p_t(\theta_0, A_t)}[L_t|X_t, A_t, \mathcal{H}_{t-1}]|X_t, \mathcal{H}_{t-1}] - \min_a \mathbb{E}_{L_t' \sim p_t(\theta_0, a)}[L_t'|X_t, a, \mathcal{H}_{t-1}]\right] \\
&= \mathbb{E}\left[\sum_{t=1}^{T} \mathbb{E}_{A_t \sim \pi_t}[\ell_t(\theta_0, A_t)|X_t, \mathcal{H}_{t-1}] - \ell_t^*(\theta_0)\right] \\
&= \mathbb{E}\left[\sum_{t=1}^{T} r_t(\pi_t, \theta_0)\right],
\end{aligned}
$$

and therefore decompose it into the following quantities,

$$
\begin{aligned}
\mathbb{E}[\mathbf{Reg}(\theta_0)] &= \mathbb{E}\left[\sum_{t=1}^{T} r_t(\pi_t, \theta_0)\right] \\
&= \mathbb{E}\left[\sum_{t=1}^{T} r_t(\pi_t, \theta_0) + \bar{r}_t(\pi_t) - \bar{r}_t(\pi_t)\right] \\
&= \mathbb{E}\left[\sum_{t=1}^{T} \overline{\mathrm{DEC}}_{\mu,t}(\pi_t) + \mu\overline{\mathrm{IG}}_t(\pi_t) + r_t(\pi_t, \theta_0) - \bar{r}_t(\pi_t)\right] \\
&= \mathbb{E}\left[\sum_{t=1}^{T} \overline{\mathrm{DEC}}_{\mu,t}(\pi_t) + \mu\overline{\mathrm{IG}}_t(\pi_t) + \mathbb{E}_{A_t \sim \pi_t}[\ell_t(\theta_0, A_t) - \bar{\ell}_t(A_t)|X_t, \mathcal{H}_{t-1}] + \bar{\ell}_t^* - \ell_t^*(\theta_0)\right] \\
&= \mathbb{E}\left[\sum_{t=1}^{T} \overline{\mathrm{DEC}}_{\mu,t}(\pi_t) + \mu\overline{\mathrm{IG}}_t(\pi_t) + \mathrm{UE}_t + \mathrm{OG}_t\right],
\end{aligned}
$$

where bounding regret would simply require bounding each of these 4 terms.