Contextual Bandits.

n arms

for $t = 1, 2, \ldots$

Nature reveals context $c_t \in \mathcal{C}$

Player chooses $a_t \in [n]$

Receives loss $\ell(c_t, a_t)$

Objective: Minimize loss $\sum_{t=1}^{T} \ell(c_t, a_t)$

First idea: ignore context and run EXP-3.

$\Rightarrow$ regret $\max_a \sum_{t=1}^{T} \ell(c_t, a_t) - \ell(c_t, a) \leq \sqrt{nT}$

Total loss $= \sum_{t=1}^{T} \ell(c_t, a_t) \leq \left(\min_a \sum_{t=1}^{T} \ell(c_t, a) \right) + \sqrt{nT}$

Second idea: Assume $|\mathcal{C}|$ is small and instantiate an independent copy of EXP-3 for each context.

$$\text{Regret} = \sum_{c \in \mathcal{C}} \max_{a \in [n]} \sum_{t : c_t = c} \left(\ell(c_t, a) - \ell(c_t, a_t) \right)$$

$$\leq \sum_{c \in \mathcal{C}} \sqrt{T_c \cdot n} = (1, \ldots, 1) \begin{pmatrix} \sqrt{T_1 n} \\ \vdots \\ \vdots \\ \sqrt{T_{|\mathcal{C}|} n} \end{pmatrix}$$

$$\leq \sqrt{|\mathcal{C}| \cdot \sum_{c \in \mathcal{C}} T_c n} = \sqrt{n \cdot |\mathcal{C}| \cdot T}$$

Total loss $\sum_{t=1}^{T} \ell(c_t, a_t) \leq \left(\sum_{c \in \mathcal{C}} \min_a \sum_{t : c_t = c} \ell(c_t, a) \right) + \sqrt{n \cdot |\mathcal{C}| \cdot T}$

We have a collection of policies Π s.t.

$\pi \in \Pi$ $\quad \pi : \mathcal{C} \to \Delta_n$ (if stochastic) $\pi : \mathcal{C} \to [n]$ (if deterministic)

Policy regret $\max_{\pi \in \Pi} \sum_{t=1}^{T} \ell(c_t, \pi) - \ell(c_t, \pi_t)$

where at each time player chooses $\pi_t \in \Pi$

then (optionally) draws $a_t \sim \pi_t(c_t)$, $\ell(c,\pi) = \mathbb{E}_{a_t \sim \pi(c_t)}[\ell(c, a_t)]$

In the case of all contexts to all actions $|\Pi| = n^{|\mathcal{C}|}$

Enumerate policies $\pi_1, \dots, \pi_{|\Pi|}$ and define

$$\ell_{t,i} = \mathbb{E}_{a \sim \pi_i(c_t)}[\ell(c_t, a)]$$

$(\pi_t \sim q_t)$

We observe $\ell(c_t, a_t)$, $a_t \sim \pi_{I_t}(c_t)$, $I_t \sim q_t$

$$\hat{\ell}_{t,i} = \frac{\pi_i(a_t | c_t)}{\sum_{k=1}^{|\Pi|} q_{t,k} \, \pi_k(a_t | c_t)} \ell(c_t, a_t)$$

$$\mathbb{E}[\hat{\ell}_{t,i}] = \sum_{j} \mathbb{P}(a_t = j \mid c_t) \frac{\pi_i(j \mid c_t)}{\sum_h q_{th}\, \pi_h(j \mid c_t)} \ell(c_t, j)$$

$$= \sum_{j} \pi_i(j \mid c_t)\, \ell(c_t, j) = \mathbb{E}_{a \sim \pi_i(c_t)} [\ell(c_t, a)]$$

$$= \ell_{t,i}$$

$$\mathbb{P}(a_t = j \mid c_t) = \sum_{s=1}^{|\pi|} \mathbb{P}(a_t = j, I_t = s \mid c_t)$$

$$= \sum_{s=1}^{|\pi|} \mathbb{P}(a_t = j \mid I_t = s, c_t)\, \mathbb{P}(I_t = s \mid c_t)$$

$$= \sum_{s=1}^{|\pi|} \pi_s(j \mid c_t) \cdot q_{t,s}$$

EXP3(γ): Exponential Weights for Exploration Exploitation

Input: Time horizon T, n arms, $\eta > 0$, $\gamma \in [0,1]$

Initialize: Player sets $p_1 = (1/n, \dots, 1/n) \in \triangle_n$. Adversary chooses $\{\ell_t\}_{t=1}^T \subset [-1,1]^n$.

for: $t = 1, \cdots, T$ — Nature reveals context c_t

Player defines $\lambda_t \in \triangle_n$ and plays $I_t \sim q_t := (1-\gamma)p_t + \gamma\lambda_t$, $a_t \sim \pi_{I_t}(c_t)$

Player suffers (and observes) loss ℓ_{t,I_t} (but does *not* observe $\ell_{t,i}$ for $i \neq I_t$)

Player computes loss estimator $\widehat{\ell}_{t,i}$ for all $i \in [n]$

Update iterates:

$$w_{t+1,i} = w_{t,i}\exp(-\eta\widehat{\ell}_{t,i}) \qquad p_{t+1,i} = w_{t+1,i} \Big/ \sum_{j=1}^n w_{t+1,j}.$$

Theorem 17. *Fix any sequence $\ell_t \in [-1,1]$ for all t. For any $\widehat{\ell}_{t,i}$ and $\eta, \gamma \geq 0$ that satisfy $\mathbb{E}[\widehat{\ell}_{t,i}|\mathcal{F}_{t-1}] = \ell_{t,i}$ and $-\eta\widehat{\ell}_{t,i} \leq 1$ for all i, t we have*

$$\max_{i=1,\dots,n} \mathbb{E}\Big[\sum_{t=1}^T \ell_{t,I_t} - \ell_{t,i}\Big] \leq 2\gamma T + \frac{\log(|\Pi|)}{\eta} + (1-\gamma)\eta\,\mathbb{E}\Big[\underbrace{\sum_{t=1}^T \sum_{j=1}^{|\Pi|} p_{t,j}\widehat{\ell}_{t,j}^2}_{\leq nT}\Big].$$

After opt. for η get $\sqrt{nT\log|\Pi|}$

$$-\eta\hat{\ell}_{t,i} \leq \eta \frac{\pi_i(a_t|c_t)}{\sum_h \pi_h(a_t|c_t)\, q_{t,h}}$$

$$\leq \eta \sum_{j=1}^n \frac{\pi_i(j|c_t)}{\sum_h \pi_h(j|c_t)\, q_{t,h}}$$

$$\leq \eta \sum_{j=1}^n \frac{\pi_i(j|c_t)}{\sum_h \pi_h(j|c_t)\,\gamma\lambda_{t,h}} \leq \frac{\eta n}{\gamma} \leq 1$$

if $\gamma = \eta n$

Claim For $P_i \in \Delta_n$ for $i = 1, \dots, m$ we have

$$\min_{\lambda\in\Delta_m} \max_{i=1,\dots,m} \sum_{j=1}^n \frac{P_{i,j}}{\sum_k \lambda_k P_{k,j}} = n$$