Exercise 7.1: Expected Payoff

Exercise 7.1: Expected Payoff

Setup:

- online learning, $k = 2$ actions
- probabilities algorithm selects each action in round i are:

$$\boldsymbol{\pi}^i = (\pi_1^i, \pi_2^i) = (2/3, 1/3)$$

- payoffs of each action in round i are:

$$\mathbf{v}^i = (v_1^i, v_2^i) = (3, 9)$$

Question: What is the expected payoff of the algorithm in round i?

(Answer on Canvas)

Lecture 7: Multi-armed Bandit Learning

Last Time:

- online learning (cont)
- warmup: geometric random variables
- follow the perturbed leader (analysis)

Lecture 7: Multi-armed Bandit Learning

Last Time:

- online learning (cont)
- warmup: geometric random variables
- follow the perturbed leader (analysis)

Today:

- multi-armed bandit learning
- reduction to online learning

Exercise 7.2: MAB-EW

Per-stage Regret Review

Recall: the per-round regret of exponential weights alg is $2h\sqrt{\ln k/n}$

- dependence on maximum value h is $O(h)$
- dependence on number of rounds n is $O(\sqrt{1/n})$
- dependence on number of actions k is $O(\sqrt{\log k})$

Exercise 7.2: MAB-EW

Setup:

- payoffs in $[0, h]$
- apply multi-armed-bandit reduction to exponential weights alg
- recall Theorem: $\mathbf{E}[\mathrm{MAB}] \geq (1 - 2\epsilon)\,\mathrm{OPT} - {}^{h\,k}/_{\epsilon^2} \ln k$
- optimally tune the learning rate ϵ for n rounds

Question: Analyze the per-round regret, what is dependence on maximum payoff h? Number of rounds n? Number of actions k? (Answer on Canvas)