CS6046 Problem Set 3

Instructor: Dr. Kota Srinivas Reddy
Jul - Nov 2023, Deadline : 29/09/2023

1. (10 marks) **Successive Elimination Algorithm (SEA) with Switching Cost**

   The performance of SEA is outlined below:

   For a given δ, with probability at least 1-δ:

   - *SEA produces the correct output*
   - *The SEA sample complexity (or total number of arm selections/pulls), which is denoted by T_{SEA} is upper bounded by $\sum_{a=1}^{K} T_a$, where*

   $$T_a = 102.\frac{\ln\left(\frac{64\sqrt{\frac{4K}{\delta}}}{\Delta_a^2}\right)}{\Delta_a^2} + 1.$$

   If we assume that each arm pull incurs a cost of Rs. 1, then we can see that with probability at least 1-δ, SEA total cost (total number of rupees required) is upper bounded by Rs. $\sum_{a=1}^{K} T_a$.

   Now, let us extend these results to the case of multi-armed bandits with arm-switching costs. Let us assume that in the multi-armed bandits with arm-switching costs, we incur a cost of Rs. 1 at time t, if the arm pulled at time t is the same as the arm pulled at time $t-1$ (i.e., no switching of arms between the time-slots $t-1$ and t or $A_{t-1} = A_t$) and we incur a cost of Rs. C at time t, if the arm pulled at time t is not the same as the arm pulled at time $t-1$ (i.e., switching of arms between the time-slots $t-1$ and t or $A_{t-1} \neq A_t$). For the multi-armed bandits with arm-switching costs, provide a reasonable upper bound on the SEA total cost.

2. (bonus - 10 marks) If someone provides (leaks) sub-optimality gaps (Δ_as), not in the same order, modify the SEA algorithm to minimize the total cost. Provide an upper bound on the total cost of the modified SEA.

   (Hint: Refer to the Successive Elimination section of the paper *Action Elimination and Stopping Conditions for the Multi-Armed Bandit and Reinforcement Learning Problems* by Eyal Even-Dar, Shie Mannor and Yishay Mansour.)

3. (10 marks) **Simple Regret Lower Bound**

   Let $\mathcal{E}$ denote the environment, in which arm rewards follow Gaussian distribution with variance 1 and means $\mu \in [0, 1]$. Show that there exists a universal constant $C > 0$ such that for all $n \geq k > 1$ and all policies π, there exists a problem instance $\nu \in \mathcal{E}$ such that

   $$\mathbb{E}[R_n^{\text{Simple}}(\pi, \nu)] \geq C\sqrt{K/n}.$$

   [Hint: Similar to cumulative regret lower bound proof.]

4. (Practice) Let $\alpha(n) = \sqrt{\frac{2\ln(4Kn^2/\delta)}{n}}$ and $n^* = \inf\{n : \alpha(n') \leq \frac{\Delta}{4} \;\; \forall n' \geq n\}$. Then, show that $n^* \leq 102.\frac{\ln\left(\frac{64\sqrt{\frac{4K}{\delta}}}{\Delta^2}\right)}{\Delta^2} + 1$.

   (Hint: Refer to Lemma 6 in the paper https://arxiv.org/abs/2005.14425.)