CS 747 (Autumn 2021): Weekly Quizzes

Instructor: Shivaram Kalyanakrishnan

August 12, 2021

Note. Provide justifications/calculations/steps along with each answer to illustrate how you arrived at the answer. You will not receive credit for giving an answer without sufficient explanation.

Submission. Write down your answer by hand, then scan and upload to Moodle. Write clearly and legibly. Be sure to mention your roll number.

Week 2

Question. Since the UCB algorithm achieves logarithmic regret on every bandit instance, we may infer that it satisfies the GLIE conditions. In this question, you are to argue from first principles that indeed UCB performs an infinite amount of exploration. To simplify our argument, we only consider a 2-armed bandit instance with arms 1 and 2. Suppose that the algorithm is (1) initialised by pulling each arm once, and (2) thereafter it is greedy with respect to the arms' upper confidence bounds at each time step, (3) breaking ties uniformly at random.

Adopting the usual notation, let u_a^t and $\hat{p}_a^t$ denote the number of pulls and the empirical mean of arm $a \in \{1, 2\}$ after $t \geq 2$ pulls (which ensures that the empirical means are well-defined). We consider an arbitrary t-length history h, summarised by t, u_1^t, $\hat{p}_1^t$, u_2^t, $\hat{p}_2^t$. We contemplate: *is it possible that one of the arms will never get pulled after encountering h?* Your task is to show that on the contrary, there exists a finite integer T (which can be defined in terms of t, u_1^t, $\hat{p}_1^t$, u_2^t, $\hat{p}_2^t$ or some subset of them) such that the T pulls following h are *guaranteed* to have at least one pull of each arm. It is okay of you unable to work out an explicit formula for T, but are still able to formally argue for its existence. Support your claims with rigorous justification, rather than appealing to "intuition" and informal observations. [4 marks]

Week 1

Question. Consider the family of n-armed bandit instances, $n \geq 2$, in which each arm $a \in \{1, 2, \ldots, n\}$ generates a 1-reward with probability p_a and a 0-reward with probability $1 - p_a$. Thus, each instance of the family is fixed by a vector $(p_1, p_2, \ldots, p_n)$, where $p_a \in [0, 1]$ for $a \in \{1, 2, \ldots, n\}$.

A round-robin algorithm undertakes $m \geq 2$ passes over the set of arms; the sequence of pulls $1, 2, \ldots, n$ is repeated m times. For each arm $a \in \{1, 2, \ldots, n\}$, let s_a denote the number of 1-rewards (interpreted as "successes") from its m pulls, and let f_a denote the number of 0-rewards (interpreted as "failures") from its m pulls (hence $s_a + f_a = m$).

a. For a fixed bandit instance $(p_1, p_2, \ldots, p_n)$, what is the probability that $s_1 = s_2 = \cdots = s_n$? Give your answer in terms of $p_1, p_2, \ldots, p_n$, and m. [2 marks]

b. Denote the total number of successes after the m passes $S = s_1 + s_2 + \cdots + s_n$. What are the mean and variance of S? Again, your answer must be in terms of $p_1, p_2, \ldots, p_n$, and m. [2 marks]

It will help to view the reward given by each pull as a random variable, noting that it is independent of the $(nm - 1)$ others. This view can facilitate an easy computation of the variance of S in part b—in your answer, be sure to explain why.

Solution.

a. Each arm a is pulled m times. The probability that it gets s_a successes and f_a failures for $0 \leq s_a \leq m$, $s_a + f_a = m$, is $\binom{m}{s_a}(p_a)^{s_a}(1 - p_a)^{f_a}$. For any fixed number of successes $s \in \{0, 1, \ldots, m\}$, the probability that all n arms get s successes is

$$\prod_{a \in \{1,2,\ldots,n\}} \binom{m}{s} (p_a)^s (1 - p_a)^{m-s}.$$

The required probability takes into account all possible values of s, and is thus

$$\sum_{s=0}^{m} \prod_{a \in \{1,2,\ldots,n\}} \binom{m}{s} (p_a)^s (1 - p_a)^{m-s}.$$

b. S is seen to be the sum of nm Bernoulli variables $X_{a,l}$ for arm $a \in \{1, 2, \ldots, n\}$ and pass $l \in \{1, 2, \ldots m\}$. The mean of $X_{a,l}$ is p_a, and its variance is $p_a(1 - p_a)$. We use

$$\mathbb{E}[S] = \sum_{a=1}^{n} \sum_{l=1}^{m} \mathbb{E}[X_{a,l}] = m \sum_{a=1}^{n} p_a.$$

Since the variables are independent, we also have

$$\mathrm{Var}[S] = \sum_{a=1}^{n} \sum_{l=1}^{m} \mathrm{Var}[X_{a,l}] = m \sum_{a=1}^{n} p_a(1 - p_a).$$

Note that if random variables X and Y are *not* independent, it is not necessary that they satisfy $\mathrm{Var}[X + Y] = \mathrm{Var}[X] + \mathrm{Var}[Y]$. In typical bandit algorithms (such as ϵ-greedy sampling), the *arm* that is pulled at some fixed time step could itself be random, disallowing the decomposition of S into $\sum_{a=1}^{n} \sum_{l=1}^{m} X_{a,l}$, which makes our variance-calculation convenient.