Multi-player Multi-Armed Bandits: Combinatorial and Decentralized Settings

Rahul Jain
USC, LA

Multi-Armed bandits are an elegant model of learning in an unknown and uncertain environment. Such models are relevant in many scenarios, and of late have received increased attention recently due to various problems of distributed control that have arisen in wireless networks, pricing models on the internet, etc. We consider a non-Bayesian multi-armed bandit setting proposed by Lai & Robbins in mid-80s. There are multiiple arms each of which generates an i.i.d. reward from an unknown distribution. There are multiple players, each of whom is choosing which arm to play. If two or more players choose the same arm, they all get zero rewards. The problem is to design a learning algorithm to be used by the players that results in an orthogonal matching of players to arms, and moreover minimizes the expected regret.

We first consider this as a bipartite matching problem. We model this combinatorial problem as a classic multi-armed bandit problem but with dependent arms, and propose an index-based learning algorithm that achieves logarithmic regret. From prior results, it is known that this is order-optimal. We then consider the distributed problem where players do not communicate or coordinate in any way. We propose a index-based algorithm that uses Bertsekas' auction mechanism to determine the bipartite matching. We show that the algorithm has expected regret at most log-cubed. This is the first distributed multi-armed bandit learning algorithm in a general setting.