“Multi-Armed Bandits in Dynamic Environments and Heavy-Tailed Rewards”

The Multi-Armed Bandit (MAB) problem is a theoretical framework that models a broad range of sequential decision-making problems. Over the last two decades, the MAB framework has experienced a rapid growth in interest from both the theoretical research community and practitioners: on one hand, there are several technical challenges posed by proving theoretical guarantees on algorithms; on the other hand, MAB algorithms have been used in relevant real-world applications such as dynamic pricing and digital advertising. In this work, we examine some of the most significant streams of research in MABs, primarily from a theoretical perspective. In particular, we aim to relax some of the core assumptions of the MAB framework, making it more suited for real-world scenarios, and provide algorithms with provable theoretical guarantees. We mainly focus on the regret minimization problem, where the decision-maker observes a realization of the reward of an action after having chosen it, and aims at maximizing the overall total reward at the end. We aim at bridging the MAB problem with Markov Decision Process (MDP) problem, and to provide MAB-style algorithm with provable theoretical guarantees in the latter. Such settings do not allow for trivial characterizations of the optimal policy, allowing past actions to affect the present. The largest literature on regret minimization in MABs deals with stochastic realizations that come from fixed and well-behaved (e.g., bounded support or sub-Gaussian) probability distributions. In this thesis, we address the heavy-tailed bandit problem, where assumptions on the reward-generating distributions are reduced to the bare minimum, and the variance may be infinite.