Cooperative and Stochastic Multi-Player Multi-Armed Bandit: Optimal Regret With Neither Communication Nor Collisions

Sébastien Bubeck
Microsoft Research

Thomas Budzinski
UBC

Mark Sellke
Stanford University

Abstract
We consider the cooperative multi-player version of the stochastic multi-armed bandit problem. The multi-player bandit problem with limited communication was first introduced roughly at the same time in Lai et al. (2008); Liu and Zhao (2010); Anandkumar et al. (2011), and has been extensively studied since then Avner and Mannor (2014); Rosenski et al. (2016); Bonnefoi et al. (2017); Lugosi and Mehrabian (2018); Boursier and Perchet (2019); Alatur et al. (2019); Bubeck et al. (2020), with various assumptions on the communication/collisions.

We study the regime where the players cannot communicate but have access to shared randomness. Previously in Bubeck and Budzinski (2020) a strategy for this regime was constructed for two players and three arms, with regret $\tilde{O}(\sqrt{T})$, and with no collisions at all between the players (with very high probability). In this paper we show that these properties (near-optimal regret and no collisions at all) are achievable for any number of players and arms. The previous strategy heavily relied on a 2-dimensional geometric intuition that was difficult to generalize in higher dimensions. We replace it by a tree-based space partition that applies in full generality. At a high level, our partitioning scheme ensures that players do not collide whenever their parameter estimates are close to each other, while allowing them to play optimally except on thin, random slices of the parameter space.

Keywords: Multi-armed bandit, multi-agent learning

Acknowledgments

The third author acknowledges funding from NSF and Stanford graduate fellowships.
References


