The Pareto Frontier of Instance-Dependent Guarantees in Multi-Player Multi-Armed Bandits with no Communication (Extended Abstract)

Allen Liu CLIU568@MIT.EDU
Massachusetts Institute of Technology

Mark Sellke MSELLKE@STANFORD.EDU
Stanford University

Editors: Po-Ling Loh and Maxim Raginsky

Abstract

We study the stochastic multi-player multi-armed bandit problem. In this problem, there are m players and $K > m$ arms and the players cooperate to maximize their total reward. However the players cannot communicate and are penalized (e.g. receive no reward) if they pull the same arm at the same time. We ask whether it is possible to obtain optimal instance-dependent regret $\widetilde{O}(1/\Delta)$ where Δ is the gap between the m-th and $m+1$-st best arms. Such guarantees were recently achieved by Pacchiano et al. (2021); Huang et al. (2022) in a model in which the players are able to implicitly communicate through intentional collisions.

Surprisingly, we show that with no communication at all, such guarantees are not achievable. In fact, obtaining the optimal $\widetilde{O}(1/\Delta)$ regret for some values of Δ necessarily implies strictly sub-optimal regret for other values. Our main result is a complete characterization of the Pareto optimal instance-dependent trade-offs that are possible with no communication. Our algorithm generalizes that of Bubeck et al. (2021). As there, our algorithm succeeds even when feedback upon collision can be corrupted by an adaptive adversary, thanks to a strong no-collision property. Our lower bound is based on topological obstructions at multiple scales and is completely new.[1]

Keywords: multi-player bandit, distributed optimization, randomized algorithms

Acknowledgments

We thank Sébastien Bubeck for suggesting that we study the gap dependent regret, and for several helpful discussions. Part of this work was completed while both authors were at Microsoft Research. A.L was supported by an NSF Graduate Research Fellowship, a Hertz Foundation Fellowship, and NSF CAREER Award CCF-1453261 and NSF Large CCF1565235. M.S. was supported by an NSF Graduate Research Fellowship, a Stanford Graduate Fellowship, and NSF grant CCF-2006489.

References

Sébastien Bubeck, Thomas Budzinski, and Mark Sellke. Cooperative and stochastic multi-player multi-armed bandit: Optimal regret with neither communication nor collisions. In *Conference on Learning Theory*, pages 821–822. PMLR, 2021.

Wei Huang, Richard Combes, and Cindy Trinh. Towards optimal algorithms for multi-player bandits without collision sensing information. In *Conference on Learning Theory*, 2022.

Aldo Pacchiano, Peter Bartlett, and Michael I Jordan. An instance-dependent analysis for the cooperative multi-player multi-armed bandit. *arXiv preprint arXiv:2111.04873*, 2021.

1. Extended abstract. Full version appears as arXiv:2202.09653, v2.