Fiftieth Annual Allerton Conference
Allerton House, UIUC, Illinois, USA
October 1 - 5, 2012

Mean Field Equilibria of Multi Armed Bandit Games

Ramki Gummadi   Ramesh Johari   Jia Yuan Yu

Much of the classical work on algorithms for multi-armed bandits focuses on rewards that are stationary over time. By contrast, we study multiarmed bandit (MAB) games, where the rewards obtained by an agent also depend on how many other agents choose the same arm. When agents interact in this way, the overall system can no longer be analyzed through the eyes of a single agent; rather, we view the agents' interactions as a dynamic game, that we call a multiarmed bandit (MAB) game. We introduce a general model of MAB games, and study a notion of equilibrium inspired by a large system approximation known as mean field equilibrium. In such an equilibrium, the proportion of agents playing the various arms, called the population profile, is assumed stationary over time; the equilibrium requires a consistency check that this stationary profile arises from the policies chosen by the agents. Our main results are as follows.

1) *Existence of MFE for MAB games.* In the model we consider, we assume that agents play a fixed policy; for example, this may be a regret-optimal policy for the classical (stationary) MAB setting. We establish existence of MFE for this model. Note that though we fix the policy agents use, this approach seems sensible for MAB games; for example, if agents use a regret-minimizing policy (such as UCB), we can show that it is approximately optimal for an MAB game.
2) *Uniqueness and convergence.* We identify a contraction condition on the arm rewards that ensures the MFE is unique, and that starting from any initial state, agents' will converge to this MFE (in the sense that eventually the population profile becomes constant). The contraction condition requires that the agent population is sufficiently mixing and that the sensitivity of the reward function to the population profile is low enough. We demonstrate via numerical experiments that in fact this result seems to hold well outside of the regime studied analytically: in fact, for a wide range of parameter choices, dynamics are observed to converge despite violation of the contraction condition.
3) *Approximation.* Under the same contraction condition used to establish uniqueness, we show an approximation result that justifies our use of MFE. In particular, we establish that if the number of agents grows large, then the dynamics of the finite agent system converge to the dynamics of the mean field model.
4) *Congestion externalities.* By congestion externalities, we refer to the case where the reward function strictly decreases as a function of the population profile component. We establish that having congestion externalities is sufficient to guarantee uniqueness of the Mean Field Equilibrium, provided that the agents adopt a policy satisfies a *positive sensitivity* to arm rewards. This condition on the policy formalizes a natural expectation that the probability of choosing any given arm should increase strictly monotone with the reward probability associated with the arm.

Authors' addresses: R. Gummadi, Management Science and Engineering, Stanford University, e-mail: gummadi@stanford.edu; R. Johari, Management Science and Engineering, Stanford University, e-mail: ramesh.johari@stanford.edu; J.Y. Yu, IBM Research, Dublin, Ireland, e-mail: jiayuanyu@ie.ibm.com. This work was supported by the National Science Foundation under grants CMMI-0948434, CNS-0904609, CCF-0832820, and CNS-0644114. The full version of the paper can be found at: http://ssrn.com/abstract=2045842

978-1-4673-4539-2/12/$31.00 ©2012 IEEE