

Bandit problems with many arms

Jia Yuan Yu (Ecole Normale Supérieure)

March 4, 2011

We consider multiarmed bandit problems where the expected reward is unimodal over partially ordered arms. In particular, the arms may belong to a continuous interval or correspond to vertices in a graph, where the graph structure represents similarity in rewards. The unimodality assumption has an important advantage: we can determine if a given arm is optimal by sampling the possible directions around it. This property allows us to quickly and efficiently find the optimal arm and detect abrupt changes in the reward distributions. For the case of bandits on graphs, we incur a regret proportional to the maximal degree and the diameter of the graph, instead of the total number of vertices.