

REGRET ANALYSIS OF A BANDIT PROBLEM FOR NORMAL DISTRIBUTIONS WITH UNKNOWN MEANS AND VARIANCES

Junya Honda

Department of Complexity Science and Engineering,
Graduate School of Frontier Sciences, The University of Tokyo.
5-1-5 Kashiwanoha, Kashiwa-shi, Chiba 277-8561, Japan,
honda@it.k.u-tokyo.ac.jp

ABSTRACT

We consider a stochastic multiarmed bandit problem such that a reward of each arm follows a normal distribution with an unknown mean and variance. This normal distribution model is one of the most natural settings of stochastic bandit problems and many algorithms have been proposed for this model. However, although some algorithms are numerically shown to perform near the theoretical bound, it has been unknown whether they can actually achieve the theoretical regret bound or not. In this talk, we analyze the UCB algorithm considered by Burnetas and Katehakis and derive a finite-time regret bound which achieves the theoretical bound asymptotically. We also consider the extension to general exponential families.