

Risk-sensitive and risk-neutral multi-armed bandits

By Uriel G. Rothblum The Alexander Goldberg Chair in Mgt. Science
Faculty of Industrial Eng. & Management Technion

Joint work with Eric V. Denardo (Yale University, USA) and
Haechurl Park(Chung-Ang University, Korea)

Abstract:

The classic result for the multi-armed bandit is that each state of each bandit (Markov chain with rewards) has an "index" that depends only on the data for its bandit, and expected discounted income is maximized by playing at each epoch a bandit whose current state has the largest index. We provide a simple and constructive proof of this result with a cubic complexity bound. Further, our analysis extends to variants of the problem which exhibit risk-averse and risk-seeking exponential utility functions. A key idea is the use of 2-parameter "domination" in place of indices.