

MINIMIZING RISK MEASURES IN BANDIT PROBLEMS

J. Y. Yu, IBM Research–Ireland, Damastown, Dublin 15, Ireland, jiyuanyu@ie.ibm.com

Bandit problems model sequential decision problems where not only are the outcomes random, but their distributions are unknown. Most of the existing work on bandit problems assumes a risk-neutral agent, whose objective is maximizing the expected reward. However, many applications have risk-averse objectives, such as minimizing the probability of failure. These problems can be modeled as minimizing one of various risk measures associated with a decision policy. In this work, we consider widely used risk measures, such as the value-at-risk and the expected shortfall, and present an efficient decision policy that estimates the risk measures from the empirical distribution of observed samples. We show a performance guarantee for this policy and apply it to an active learning problem where we selectively seek feedback from human experts.