The Finite-Horizon Two-Armed Bandit Problem with Binary Responses: A Multidisciplinary Survey of the History, State of the Art, and Myths

Peter Jacko
Lancaster University, UK

Abstract

We consider the two-armed bandit problem, which often naturally appears per se or as a subproblem in some multi-armed generalizations, and serves as a starting point for introducing additional problem features. The consideration of binary responses is motivated by its widespread applicability and by being one of the most studied settings. We focus on the undiscounted finite-horizon objective, which is the most relevant in many applications. We make an attempt to unify the terminology as this is different across disciplines that have considered this problem, and present a unified model cast in the Markov decision process framework, with subject responses modelled using the Bernoulli distribution, and the corresponding Beta distribution for Bayesian updating. We give an extensive account of the history and state of the art of approaches from several disciplines, including design of experiments, Bayesian decision theory, naive designs, reinforcement learning, biostatistics, and combination designs. We evaluate these designs, together with a few newly proposed, accurately computationally (using a newly written package in Julia programming language by the author) in order to compare their performance. We show that conclusions are different for moderate horizons (typical in practice) than for small horizons (typical in academic literature reporting computational results). We further list and clarify a number of myths about this problem, e.g., we show that, computationally, much larger problems can be designed to Bayes-optimality than what is commonly believed.