Multi-Armed Bandit Problem

Sunith Suresh, Lin Xiao, Ilan Man and Sanjay Hariharan

April 23, 2016

1 Introduction

2 Multi-armed bandit review

- define problem: policy, machines, objective, regret
- exploitation vs exploration
- examples of applications

2.1 ϵ-greedy

- basic algorithm
- linear regret

2.2 upper confidence bounds

- explain concept of bounding regret
- application of Hoeffding's Inequality
- calculate UCB

3 Bayesian approach

- define rewards as bernoulli(π_i)
- select machine with probability θ_i
- update belief of machine
- goal is to maximize expecte reward

3.1 Thompson Sampling: Hueristic

- Lin

3.2 Analytical theory

- Sanjay/Lin

3.3 Dynamic programming and Gittins

- Sunith