CS 747 : Foundations of Intelligent Learning Agents
Assignment 1

Arka Sadhu - 140070011

August 11, 2017

1 Epsilon-Greedy

The epsilon-greedy algorithm works as follows :

- We choose some ε in the range $[0, 1]$.
- Then we choose the bandit with the highest empirical mean with probability $1 - \varepsilon$ and with probability ε sample an arm at random.
- ε is a constant given by the user.

2 Upper Confidence Bound (UCB)

The upper confidence bound (UCB) algorithm works as follows :

- We first pull each arm once in a round robin fashion.
- Then we compute the empirical mean of each arm. This is followed by an additional term which then gives the ucb for the corresponding arm.

$$ucb_a^t = \hat{p_a}^t + \sqrt{\frac{2 * ln(t)}{u_a^t}}$$

- At each instance we choose the arm with the highest ucb value.

3 KL-UCB

This is the KL version of the UCB and works as follows :

- We again pull each arm once in a round robin fashion.
- Then for each arm we define a parameter q_a such that $q_a \in [\hat{p_a}, 1]$ and it is the least real number to satisfy the inequality 1

$$u_a^t KL(\hat{p_a^t}, q) \geq ln(t) + cln(ln(t)) \tag{1}$$

- We then choose the arm with the highest q_a
- Since KL is a monotonically increasing function, we employ binary search algorithm to search for q_a.

4 Thompson Sampling

Thompson Sampling algorihtm works as follows:

- We start by pulling each arm once in a round robin fashion till each arm is sampled once.
- Then we note each success and failure for a particular arm. Then we generate a beta distribution whose parameters are $\alpha = success + 1$ and $\beta = failures + 1$.
- We then sample a number x from the generated beta distribution for the corresponding arm and choose the arm with highest number sampled.