Efficient Algorithms for Infinite-Armed Bandit

Arghya Roy Chaudhuri
under the guidance of
Prof. Shivaram Kalyanakrishnan

Department of Computer Science and Engineering
Indian Institute of Technology Bombay

April 1, 2016

What is a Multi Armed Bandit ?

Table: Rewards from slot machines

Machines:						
True Mean	0.9	0.5	0.6	0.7	0.1	0.7

What is a Multi Armed Bandit ?

Table: Rewards from slot machines

Machines:						
True Mean	0.9	0.5	0.6	0.7	0.1	0.7
Round 1	1	1	0	1	0	0

What is a Multi Armed Bandit ?

Table: Rewards from slot machines

Machines:						
True Mean	0.9	0.5	0.6	0.7	0.1	0.7
Round 1	1	1	0	1	0	0
Round 2	-	0	-	-	-	-

What is a Multi Armed Bandit ?

Table: Rewards from slot machines

Machines:						
True Mean	0.9	0.5	0.6	0.7	0.1	0.7
Round 1	1	1	0	1	0	0
Round 2	-	0	-	-	-	-
Round 3	1	-	-	-	-	-

What is a Multi Armed Bandit ?

Table: Rewards from slot machines

Machines:						
True Mean	0.9	0.5	0.6	0.7	0.1	0.7
Round 1	1	1	0	1	0	0
Round 2	-	0	-	-	-	-
Round 3	1	-	-	-	-	-
Round 4	1	-	-	-	-	-

Table: Rewards from slot machines

Machines:						
True Mean	0.9	0.5	0.6	0.7	0.1	0.7
Round 1	1	1	0	1	0	0
Round 2	-	0	-	-	-	-
Round 3	1	-	-	-	-	-
Round 4	1	-	-	-	-	-
Round 5	0	-	-	-	-	-

What is a Multi Armed Bandit ?

Table: Rewards from slot machines

Machines:						
True Mean	0.9	0.5	0.6	0.7	0.1	0.7
Round 1	1	1	0	1	0	0
Round 2	-	0	-	-	-	-
Round 3	1	-	-	-	-	-
Round 4	1	-	-	-	-	-
Round 5	0	-	-	-	-	-
Round 6	-	-	-	1	-	-

What is a Multi Armed Bandit ?

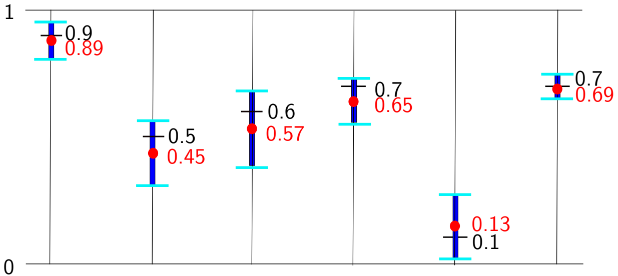
Table: Rewards from slot machines

Machines:						
True Mean	0.9	0.5	0.6	0.7	0.1	0.7
Round 1	1	1	0	1	0	0
Round 2	-	0	-	-	-	-
Round 3	1	-	-	-	-	-
Round 4	1	-	-	-	-	-
Round 5	0	-	-	-	-	-
Round 6	-	-	-	1	-	-

Objective: Output the arm with the highest expected reward with PAC guarantee incurring **minimum** number of samples

Key Principle: Confidence Bounds

Figure: Visualization of Confidence Bounds

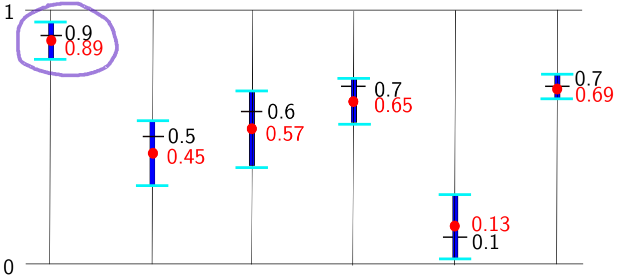


Approach:

- Track confidence bounds for each arm

Key Principle: Confidence Bounds

Figure: Visualization of Confidence Bounds

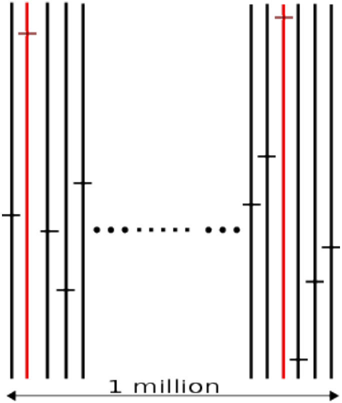


Approach:

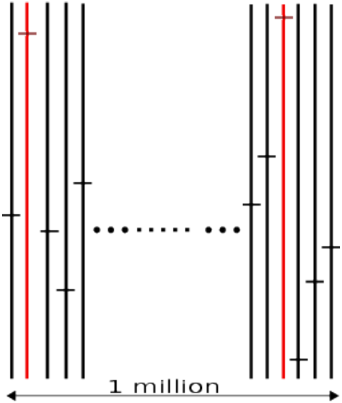
- Track confidence bounds for each arm
- Return an arm whose LCB exceeds UCB of all the other arms

Our Problem

What if the number of arms is too large ?



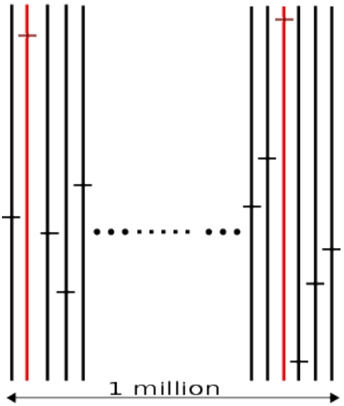
What if the number of arms is too large ?



Problem Definition: Find an arm from an infinite set of arms whose true mean reward is greater than $(1-\rho)^{\text{th}}$-quantile (for $0 < \rho < 1$) of distribution of rewards over arms.

Our Problem

What if the number of arms is too large ?



Problem Definition: Find an arm from an infinite set of arms whose true mean reward is greater than $(1-\rho)^{\text{th}}$-quantile (for $0 < \rho < 1$) of distribution of rewards over arms.

Applications:

- Action planning under uncertainty where number of actions is too huge to explore exhaustively
- Playing games where number of strategies is too high and obtained gain by applying a strategy can't be predicted beforehand.