Multi-Armed Bandits with Bounded Arm-Memory:

Near-Optimal Guarantees for Best-Arm Identification and Regret Minimization

Arnab Maiti, Indian Institute of Technology, Kharagpur

Vishakha Patil, Indian Institute of Science, Bangalore

Arindam Khan, Indian Institute of Science, Bangalore

Multi-Armed Bandits

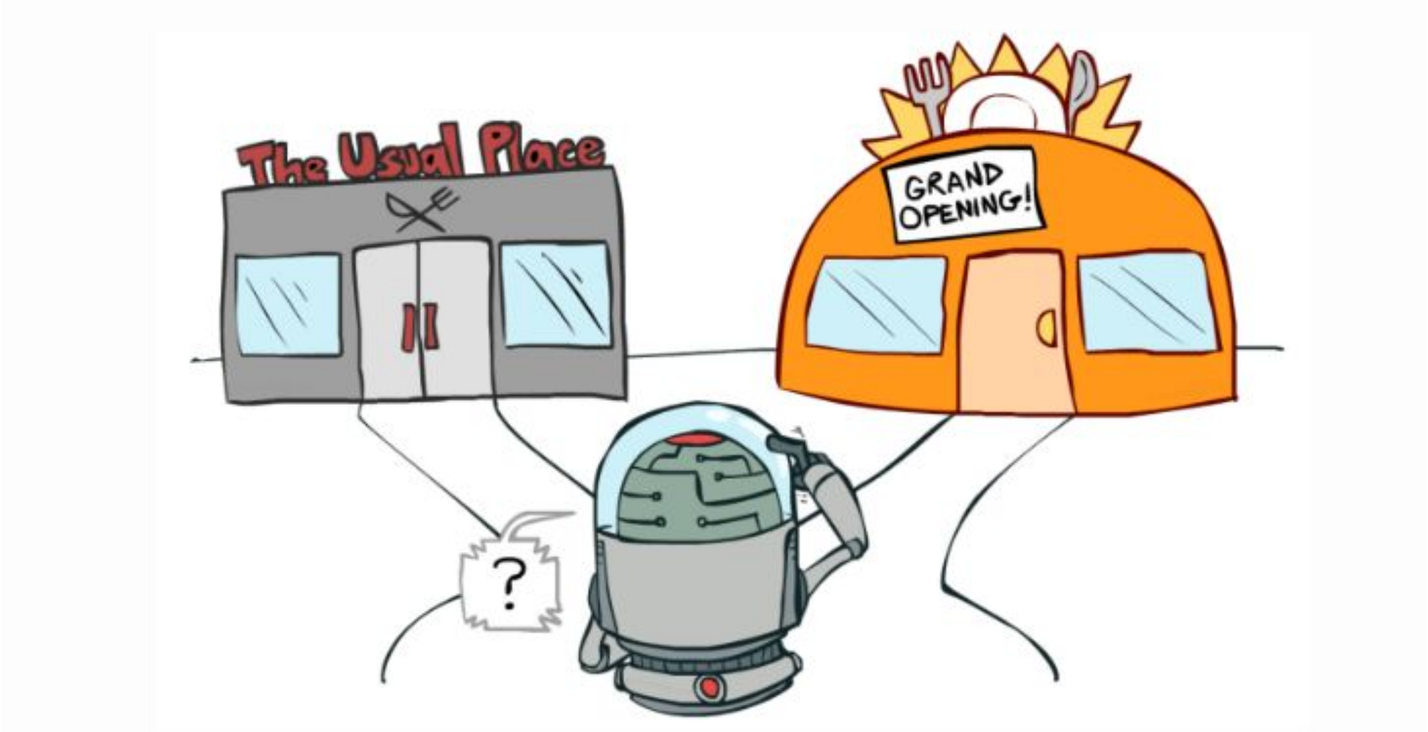


Image source: UC Berkeley AI course slide, lecture 11

Streaming Model

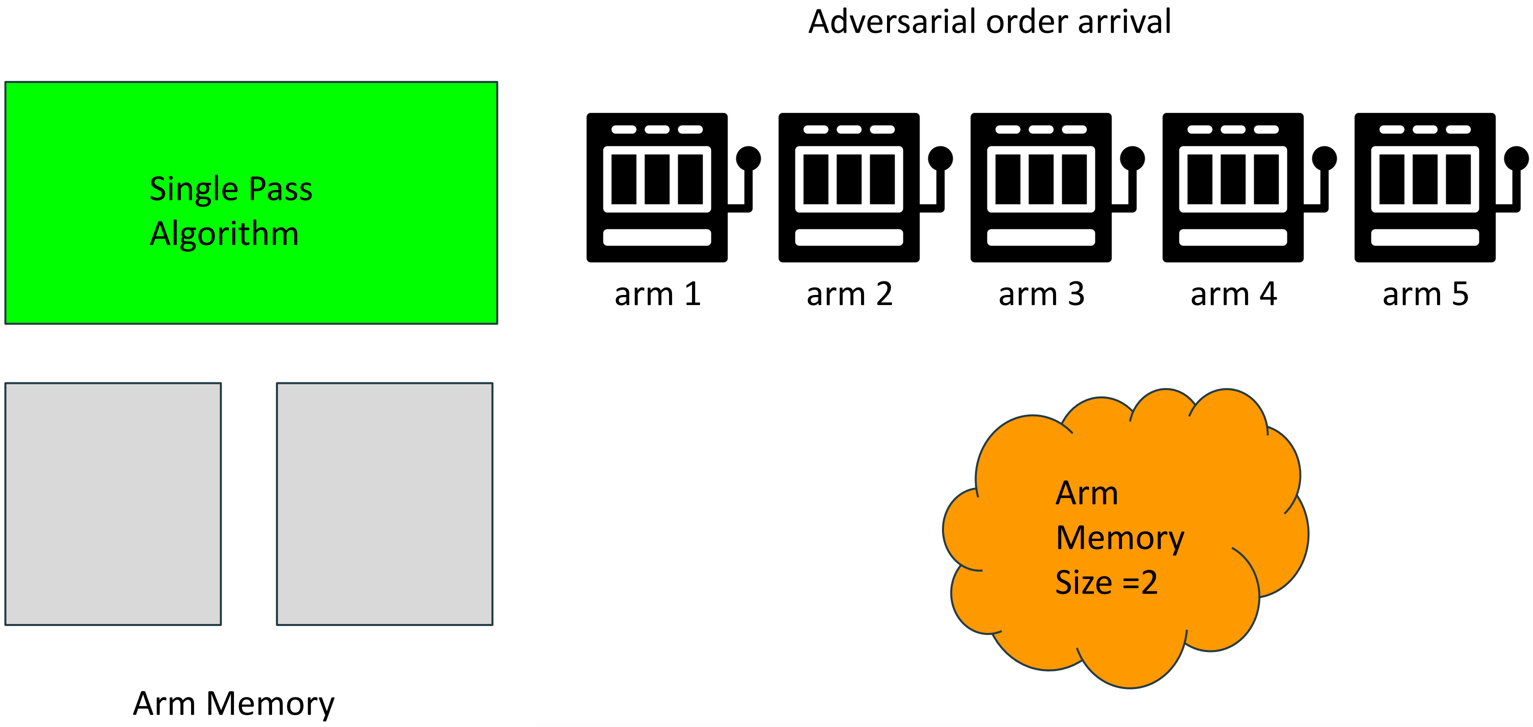


Image source: slot machine by Deemak Daksina from the Noun Project

Streaming Model

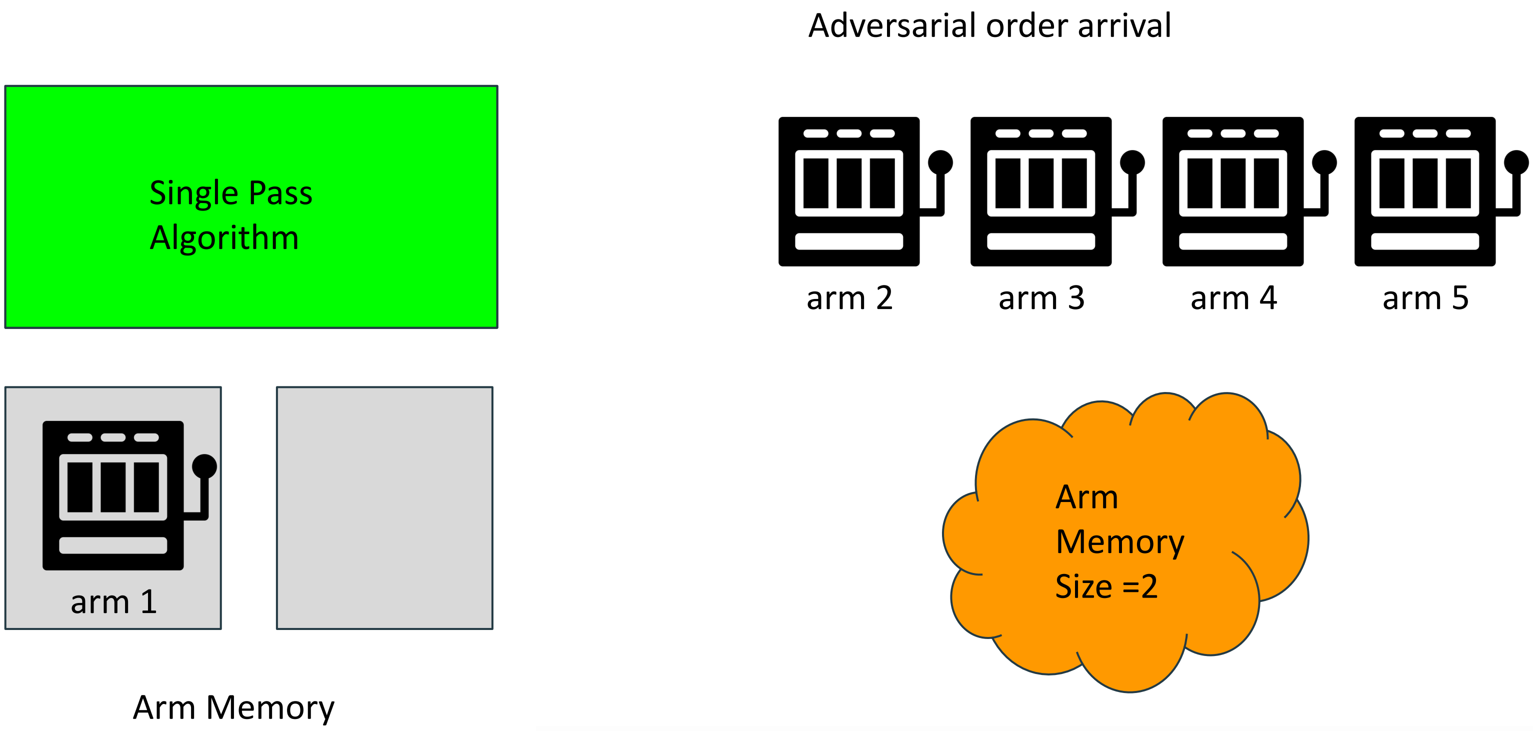


Image source: slot machine by Deemak Daksina from the Noun Project

Streaming Model

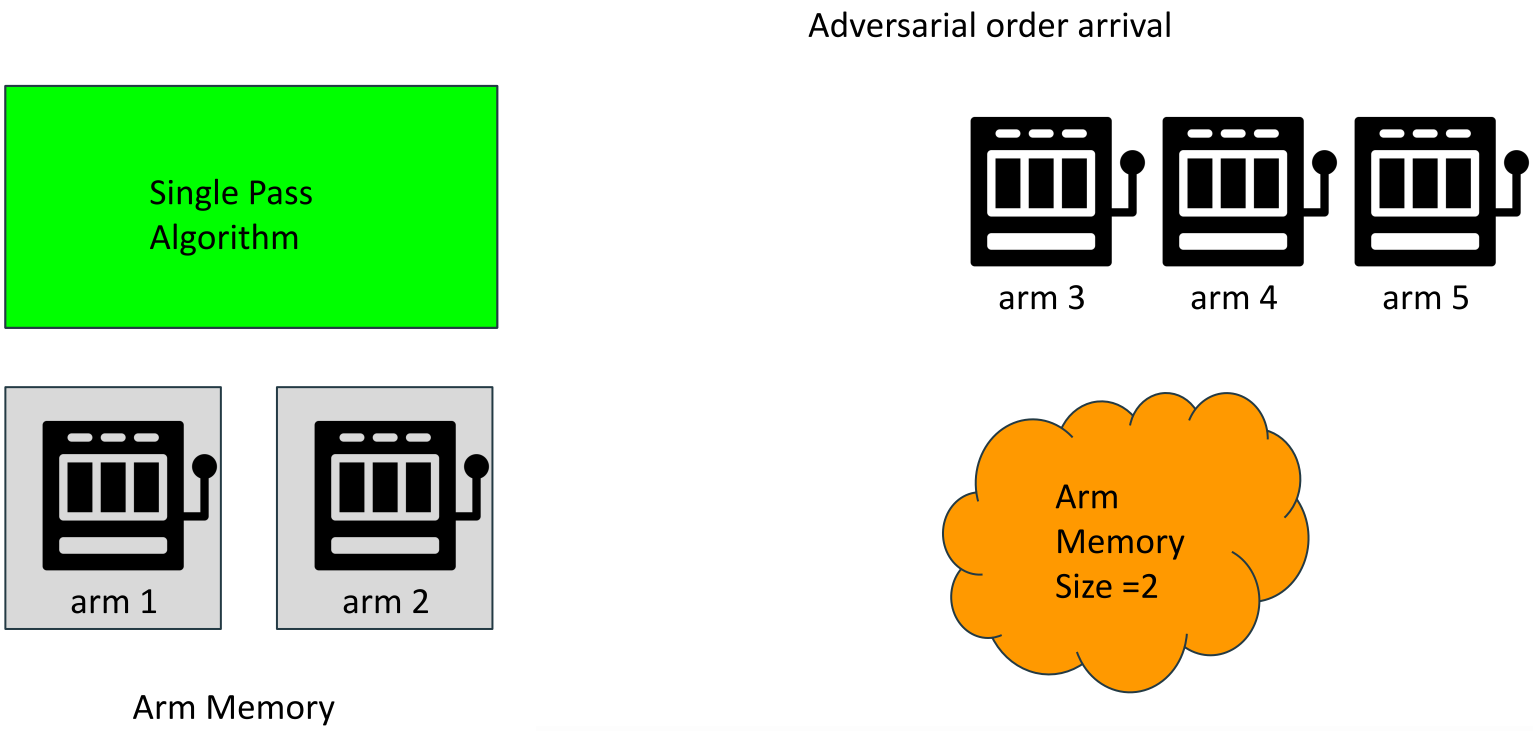


Image source: slot machine by Deemak Daksina from the Noun Project

Streaming Model

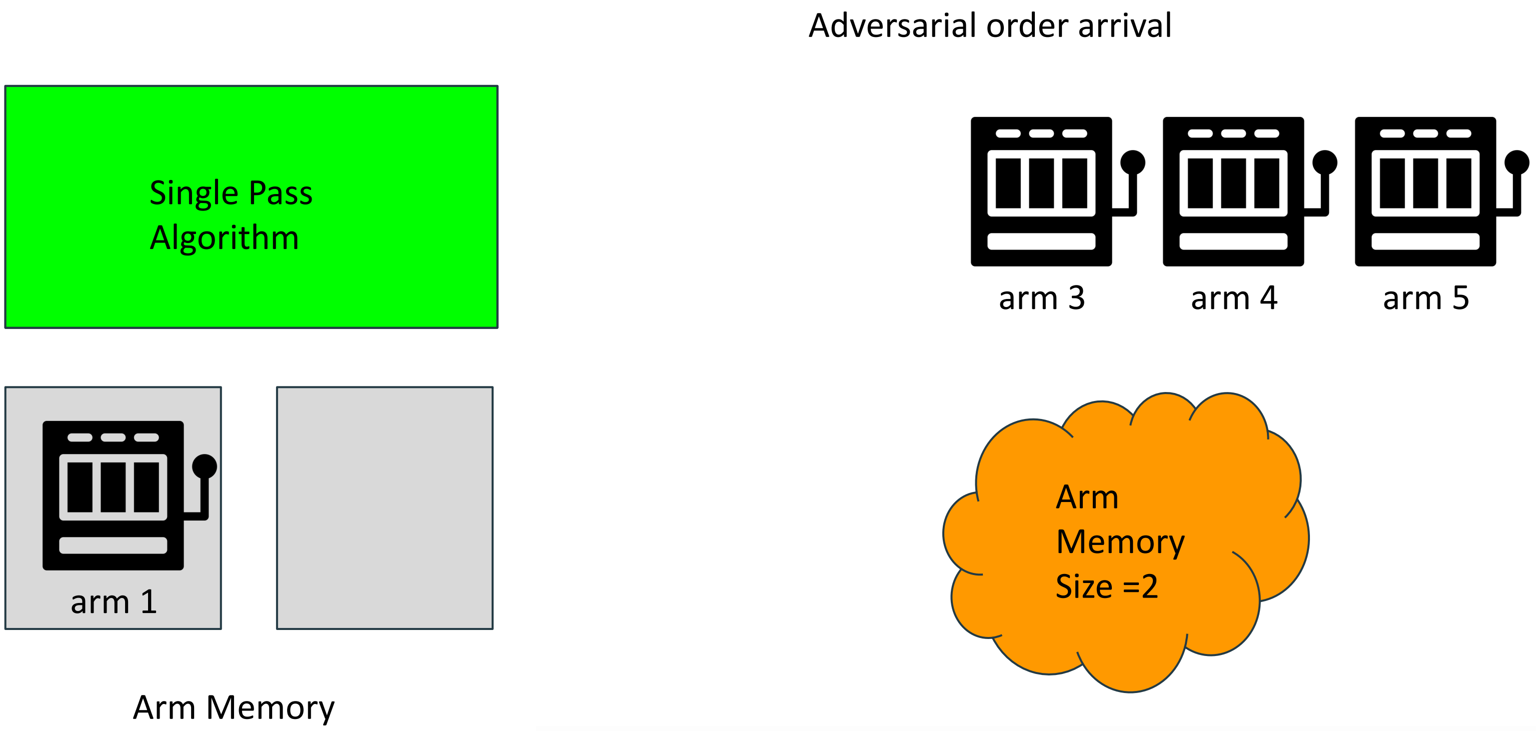


Image source: slot machine by Deemak Daksina from the Noun Project

Streaming Model

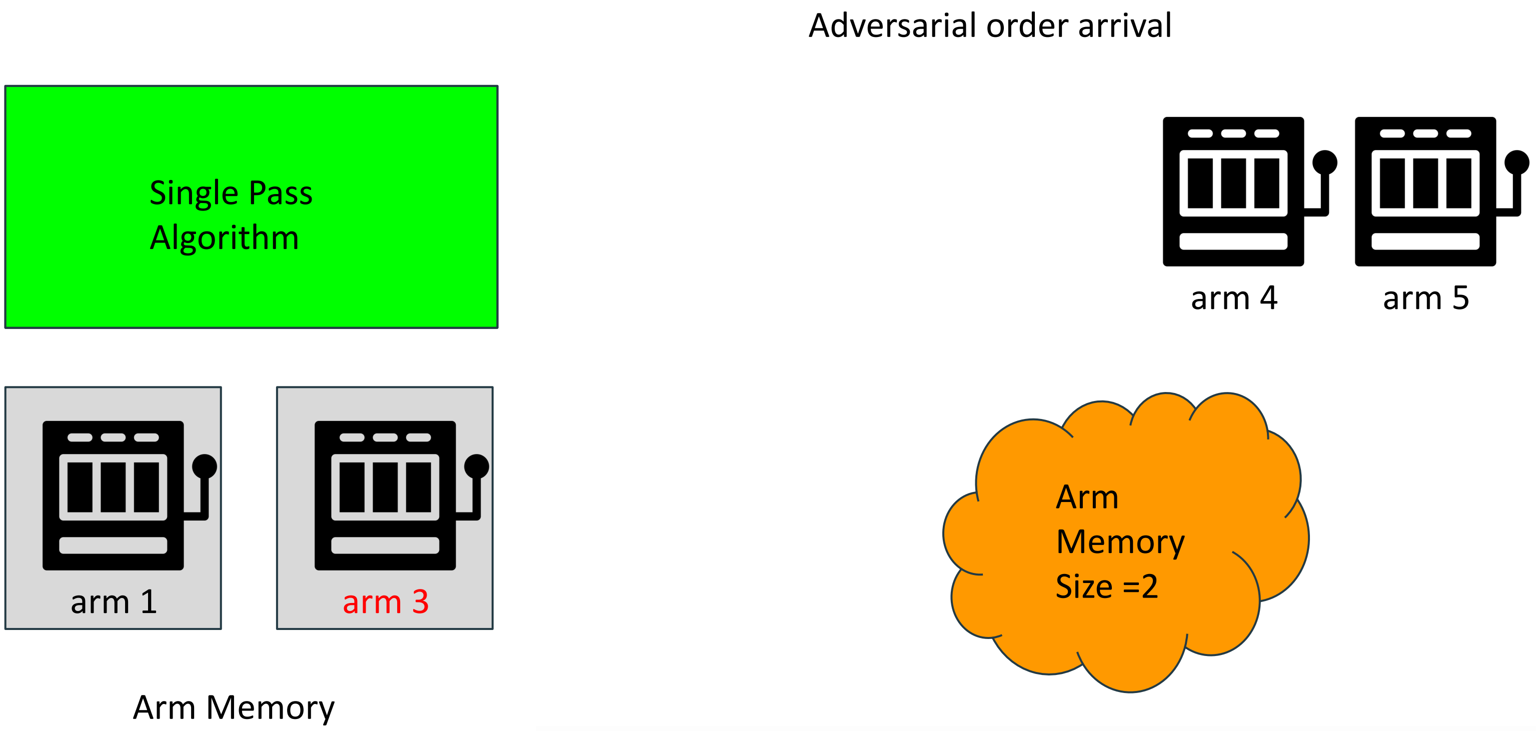


Image source: slot machine by Deemak Daksina from the Noun Project

Application: Job Interviews

Arm
Memory
Size =2

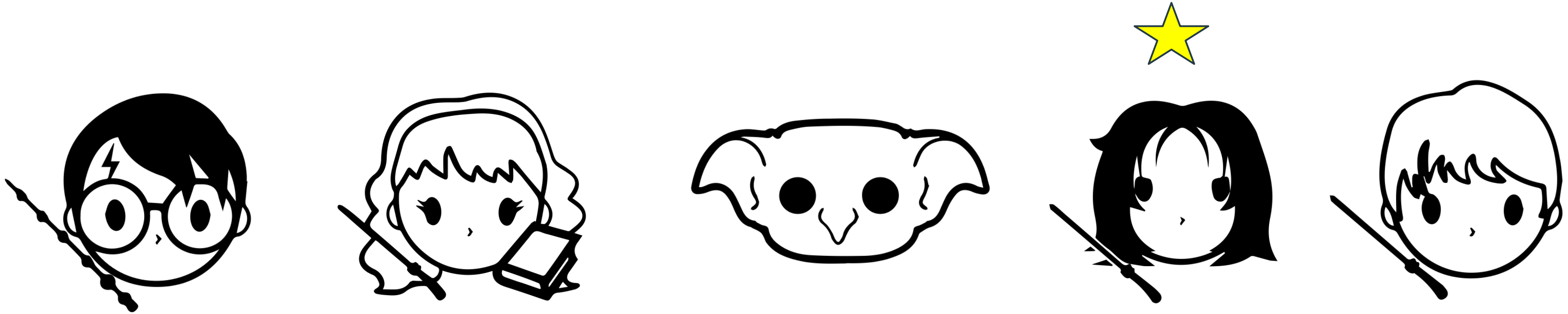


Image source: SoyGalem from the Noun Project

Application: Job Interviews

Arm
Memory
Size =2

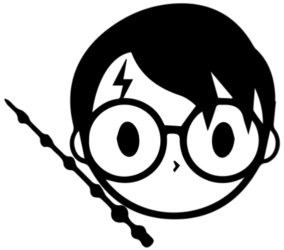


Image source: SoyGalem from the Noun Project

Application: Job Interviews

Arm
Memory
Size =2

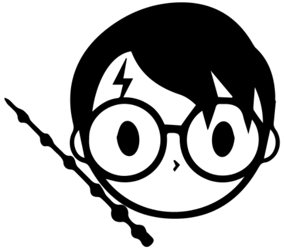


Image source: SoyGalem from the Noun Project

Application: Job Interviews

Arm
Memory
Size =2

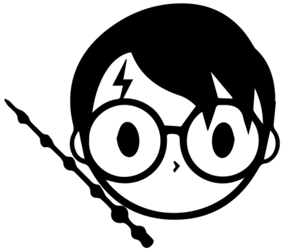


Image source: SoyGalem from the Noun Project

Application: Job Interviews

Arm
Memory
Size =2

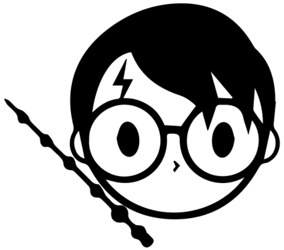
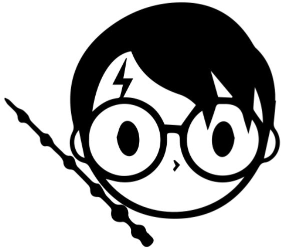


Image source: SoyGalem from the Noun Project

Application: Job Interviews

Arm
Memory
Size =2



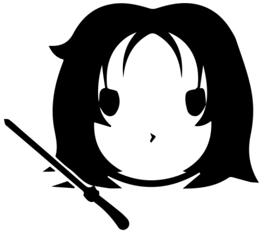


Image source: SoyGalem from the Noun Project

Application: Job Interviews

Arm
Memory
Size =2

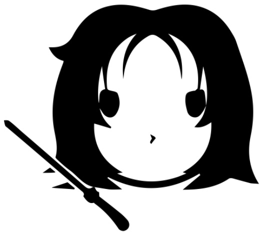
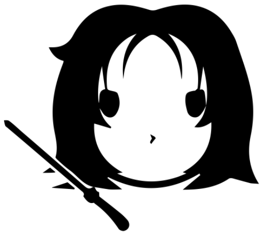


Image source: SoyGalem from the Noun Project

Application: Job Interviews

Arm
Memory
Size =2



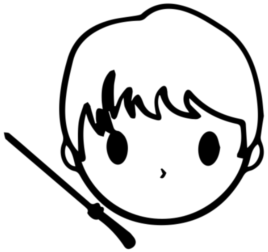


Image source: SoyGalem from the Noun Project

Application: Job Interviews

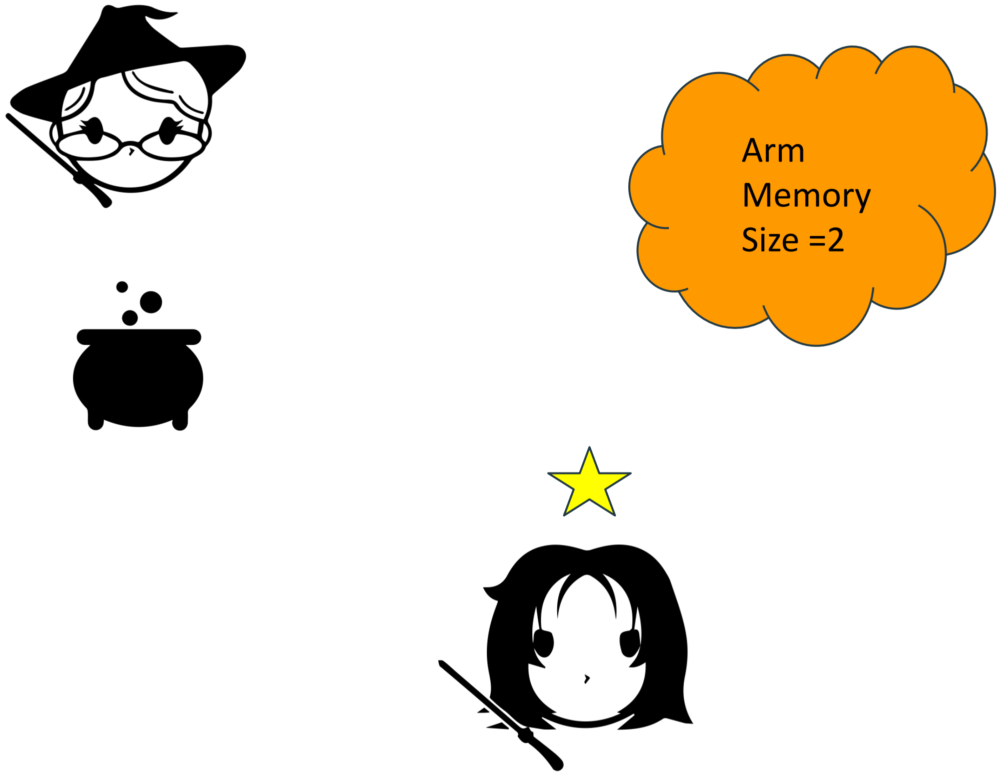


Image source: SoyGalem from the Noun Project

Regret Minimization

Stochastic MAB Instance with expected rewards

$$< \mu_1, \mu_2, \ldots, \mu_n >$$

Let $\mu^* = \max(\mu_1, \ldots, \mu_n)$

Let μ_{i_t} be the mean of the arm sampled in the time step t

Expected Cumulative Regret in T time steps

$$\mathbb{E}[R(T)] = \mu^* \cdot T - \sum_{t=1}^{T} \mathbb{E}[\mu_{i_t}]$$

Goal: Minimize $\mathbb{E}[R(T)]$

Best-Arm Identification

Stochastic MAB Instance with expected rewards

$$< \mu_1, \mu_2, \ldots, \mu_n >$$

ε-best arm is an arm with mean at least $\mu^* - \varepsilon$

(ε, δ) -PAC Algorithm: Finds an ε-best arm with probability at least $1 - \delta$

Goal: (ε, δ)-PAC algorithm such that total number of arm pulls is minimized.

Results Overview

Results Overview

Main Result 1: Any single-pass algorithm will incur at least

$$\Omega\left(\frac{n^{1/3}T^{2/3}}{m^{7/3}}\right)$$

expected cumulative regret when $m < n$.
This lower bound even holds for random order arrival.

Results Overview

Main Result 1: Any single-pass algorithm will incur at least

$$\Omega\left(\frac{n^{1/3}T^{2/3}}{m^{7/3}}\right)$$

expected cumulative regret when $m < n$.
This lower bound even holds for random order arrival.

Main Result 2: (ε, δ) - PAC algorithm with

$O(r)$ arm memory

Optimal r-round sample complexity

Results Overview

Main Result 1: Any single-pass algorithm will incur at least

$$\Omega\left(\frac{n^{1/3}T^{2/3}}{m^{7/3}}\right)$$

expected cumulative regret when $m < n$.
This lower bound even holds for random order arrival.

$$O\left(\frac{n}{\varepsilon^2}\left(\mathtt{ilog}^{(\mathtt{r})}(n) + \log\left(\frac{1}{\delta}\right)\right)\right)$$

Main Result 2: (ε, δ) - PAC algorithm with

$O(r)$ arm memory

Optimal r-round sample complexity

Results Overview

Main Result 1: Any single-pass algorithm will incur at least

$$\Omega\left(\frac{n^{1/3}T^{2/3}}{m^{7/3}}\right)$$

expected cumulative regret when $m < n$.
This lower bound even holds for random order arrival.

$$O\left(\frac{n}{\varepsilon^2}\left(\mathtt{ilog}^{(\mathtt{r})}(n) + \log\left(\frac{1}{\delta}\right)\right)\right)$$

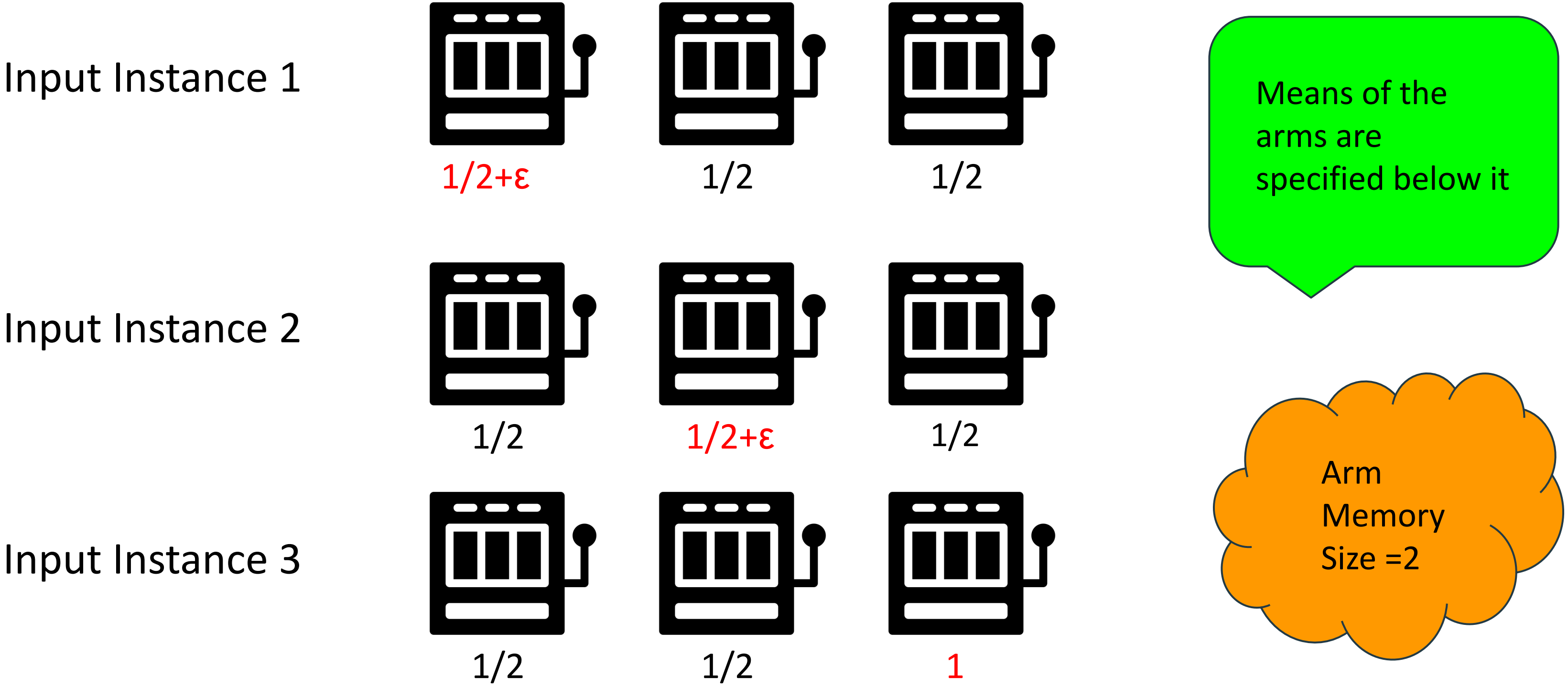
Main Result 2: (ε, δ) - PAC algorithm with

$O(r)$ arm memory

Optimal r-round sample complexity

Corollary: (ε, δ)- PAC algorithm with $O(\log^* n)$ arm memory with optimal worst-case sample complexity

Regret Minimization Lower Bound

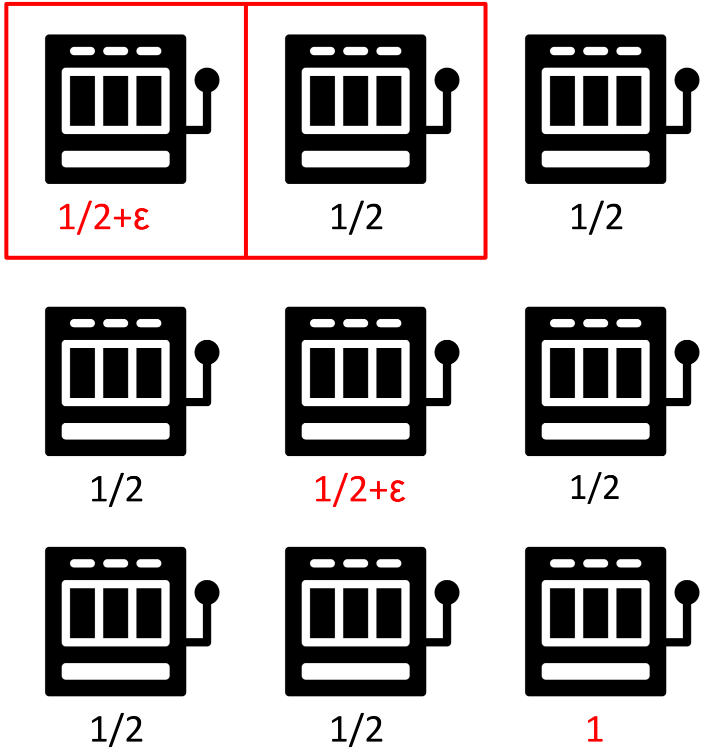


Regret Minimization Lower Bound

Input Instance 1

Input Instance 2

Input Instance 3



Means of the arms are specified below it

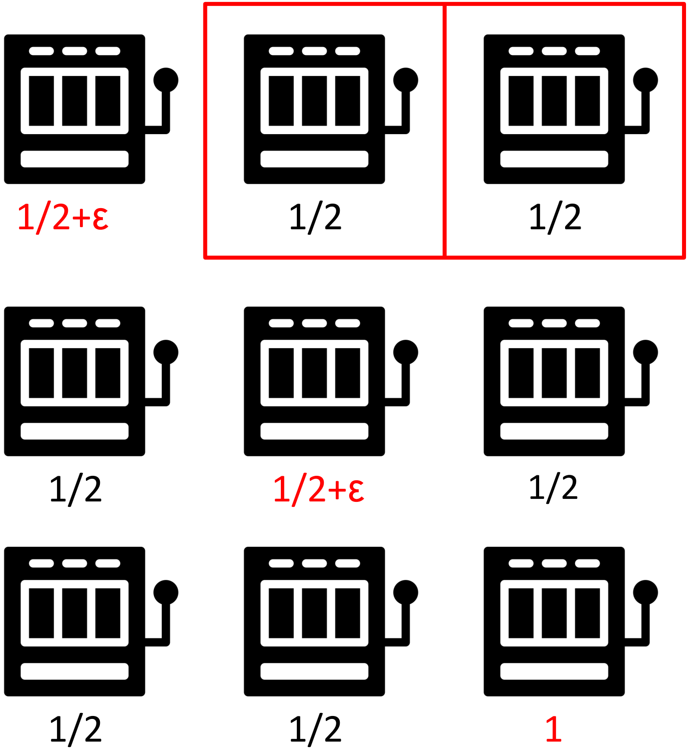
Arm Memory Size =2

Regret Minimization Lower Bound

Input Instance 1

Input Instance 2

Input Instance 3



Means of the arms are specified below it

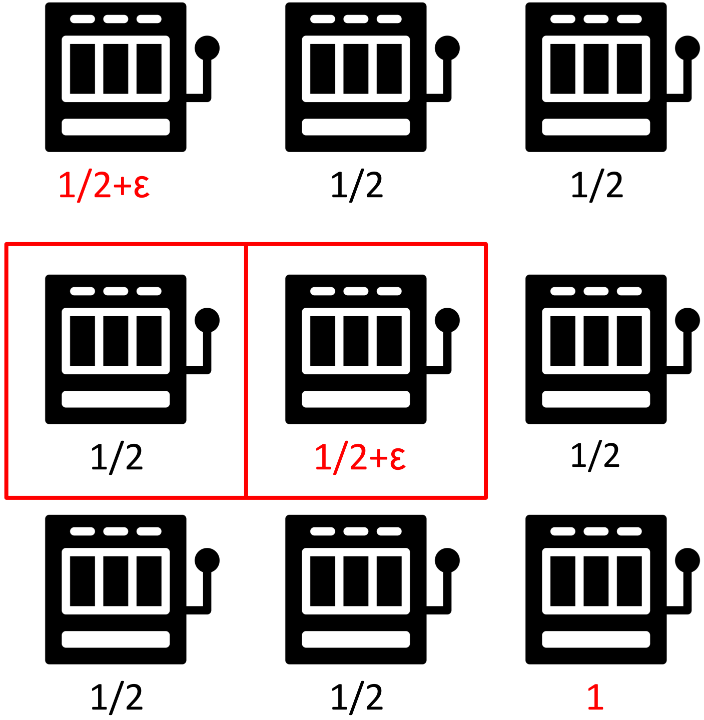
Arm Memory Size =2

Regret Minimization Lower Bound

Input Instance 1

Input Instance 2

Input Instance 3



Means of the arms are specified below it

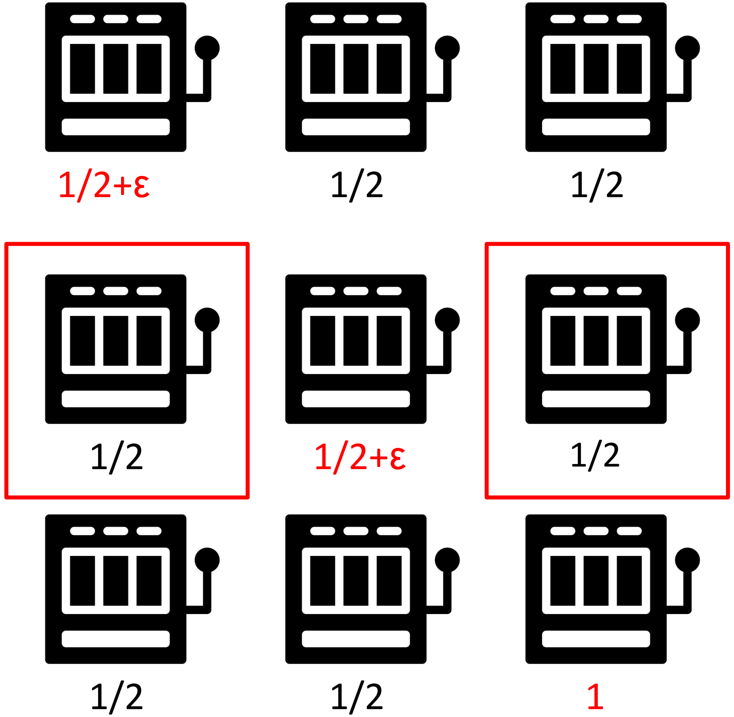
Arm Memory Size =2

Regret Minimization Lower Bound

Input Instance 1

Input Instance 2

Input Instance 3



Means of the arms are specified below it

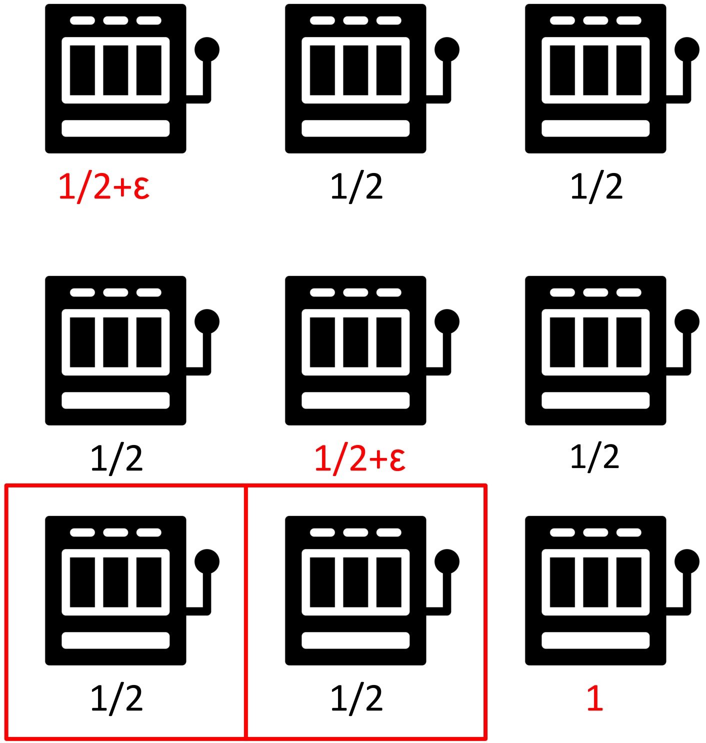
Arm Memory Size =2

Regret Minimization Lower Bound

Input Instance 1

Input Instance 2

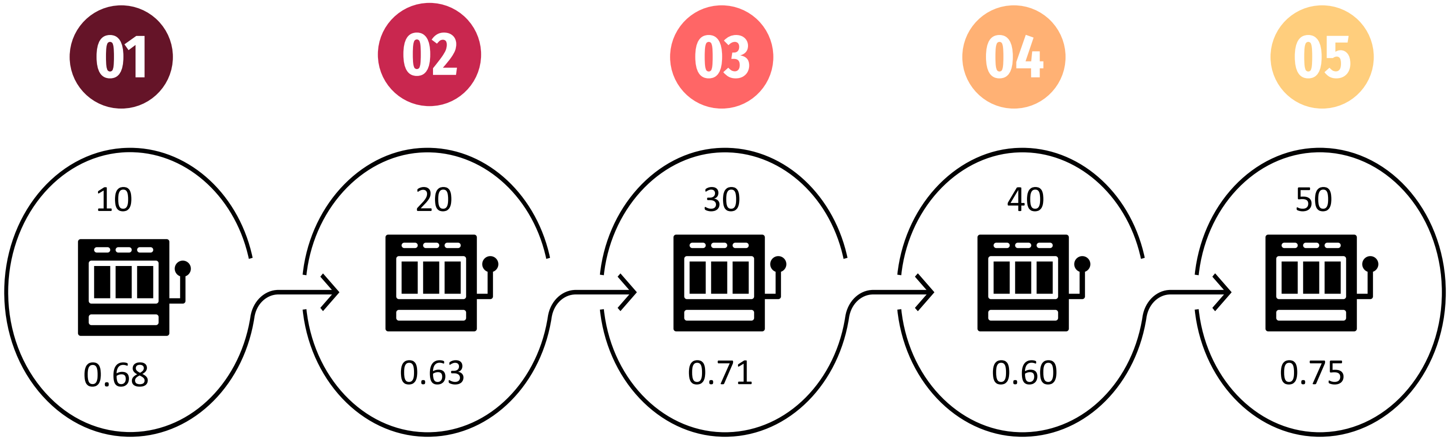
Input Instance 3



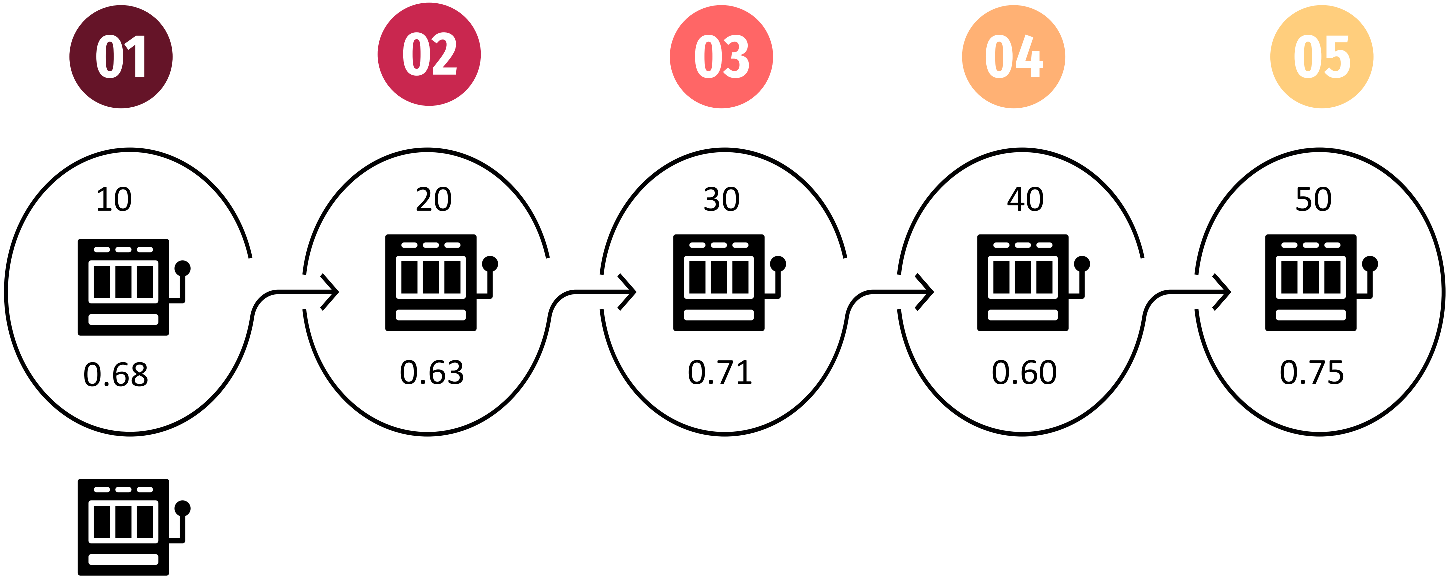
Means of the arms are specified below it

Arm Memory Size =2

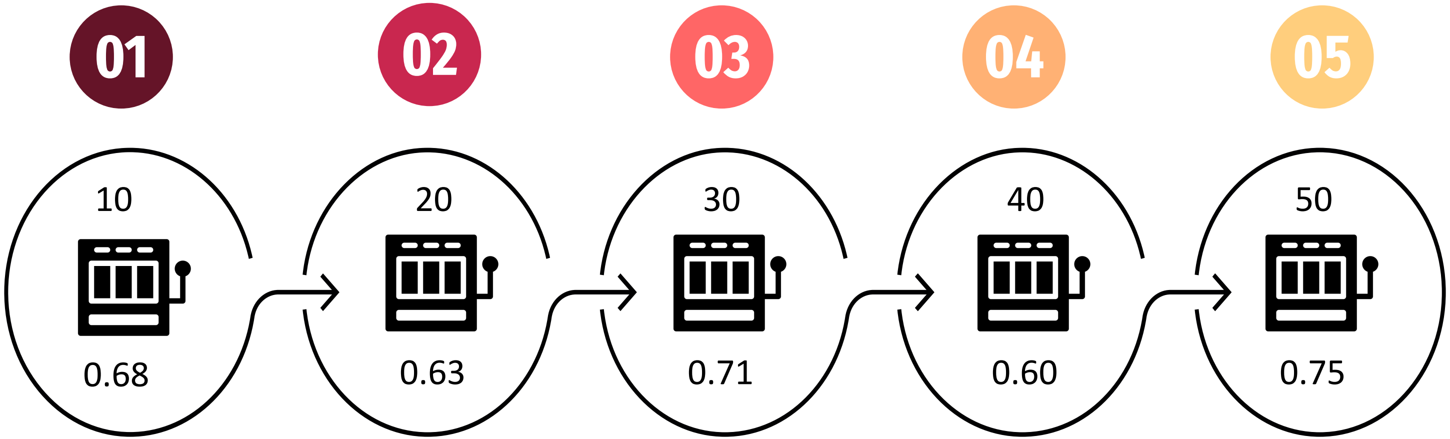
Best-Arm Identification Algorithm

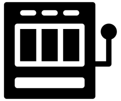


Best-Arm Identification Algorithm



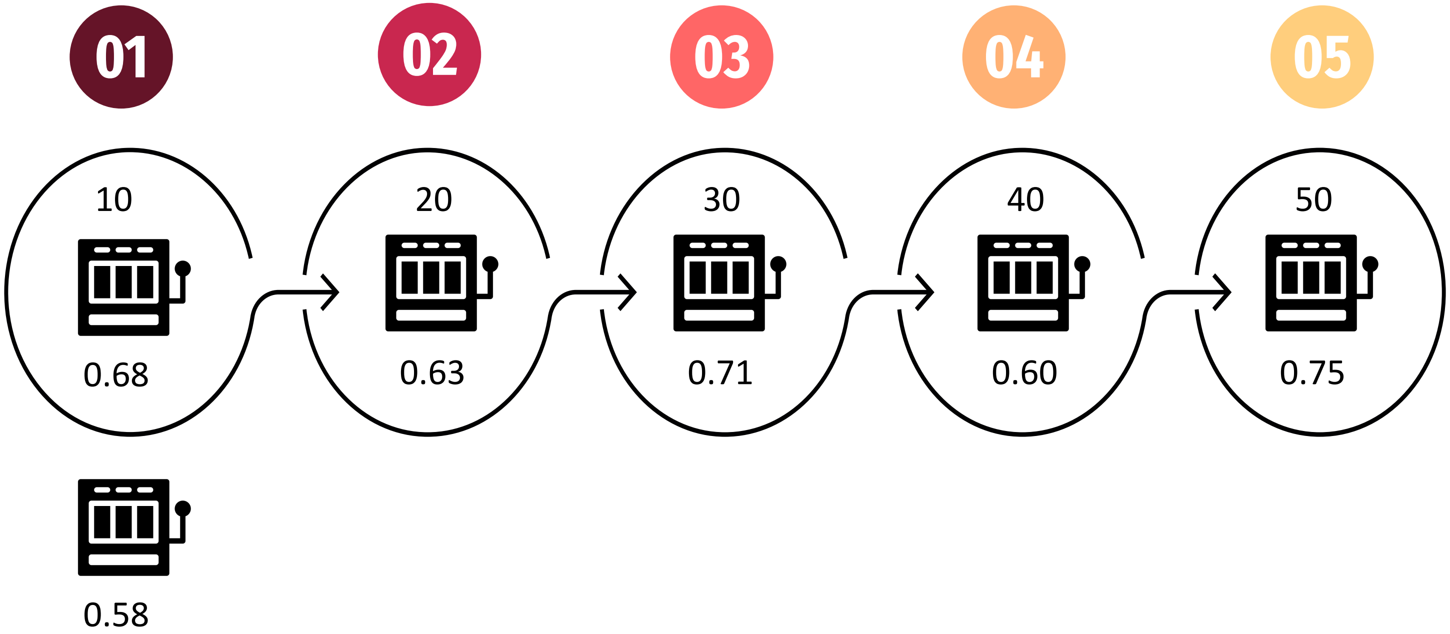
Best-Arm Identification Algorithm



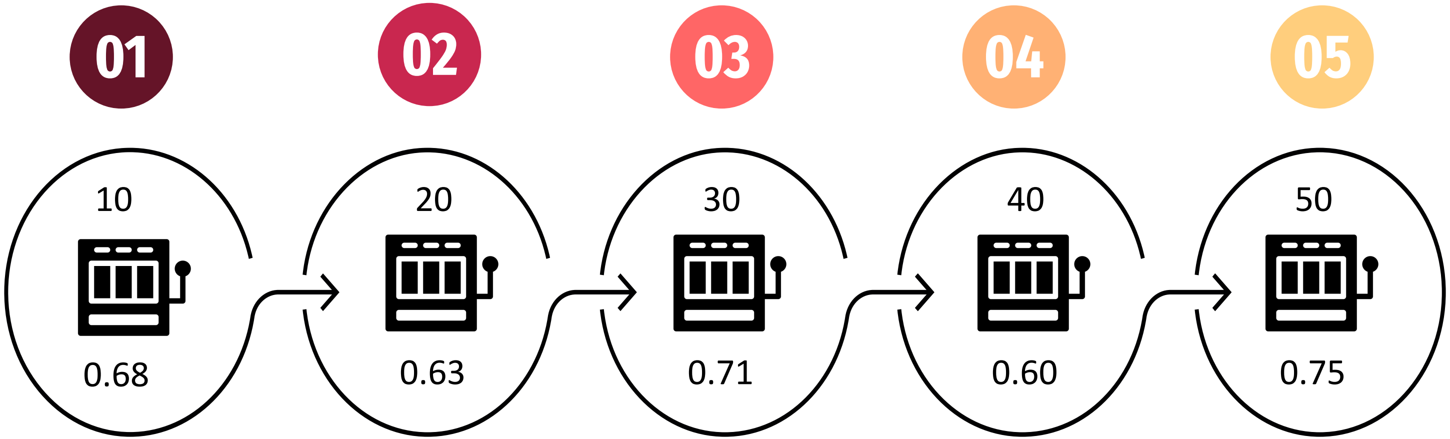


Sample for s1 times

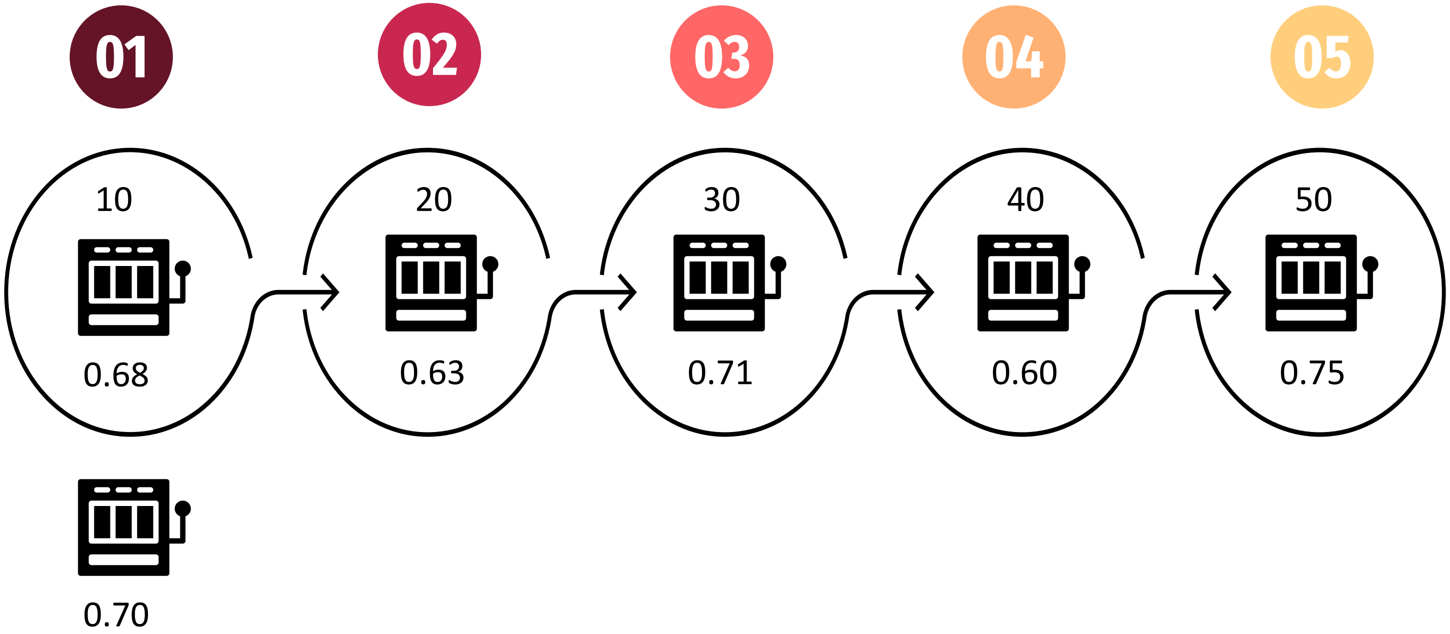
Best-Arm Identification Algorithm



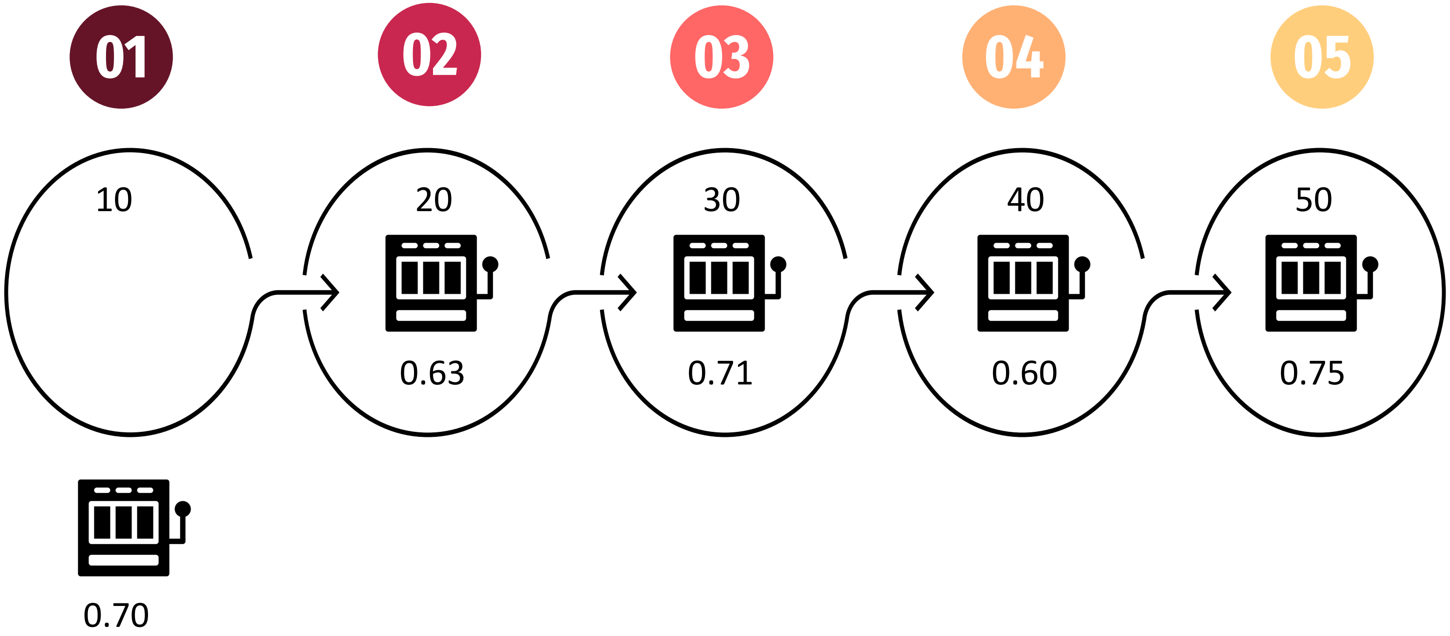
Best-Arm Identification Algorithm



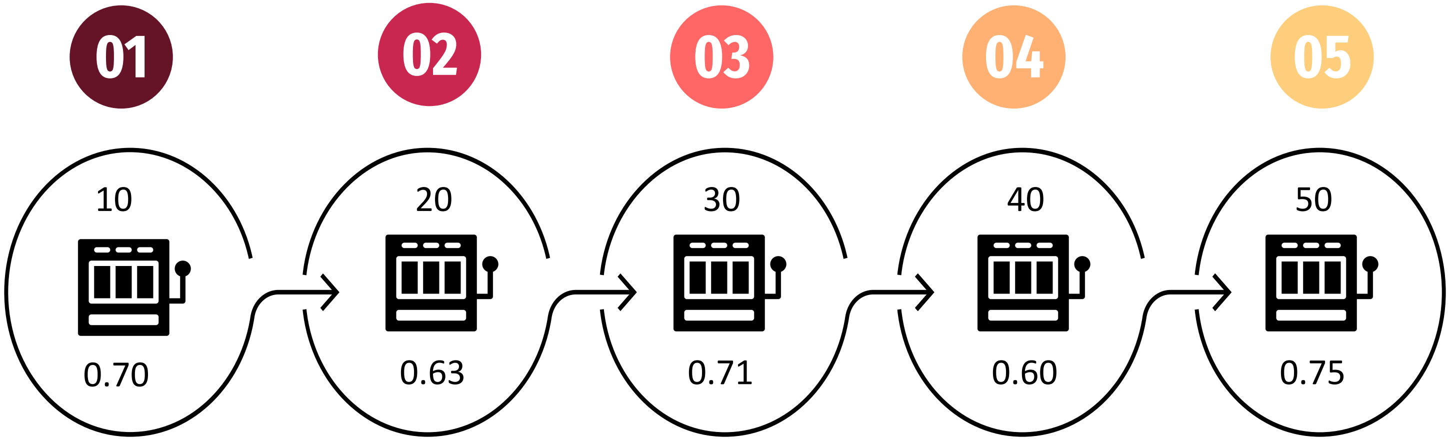
Best-Arm Identification Algorithm



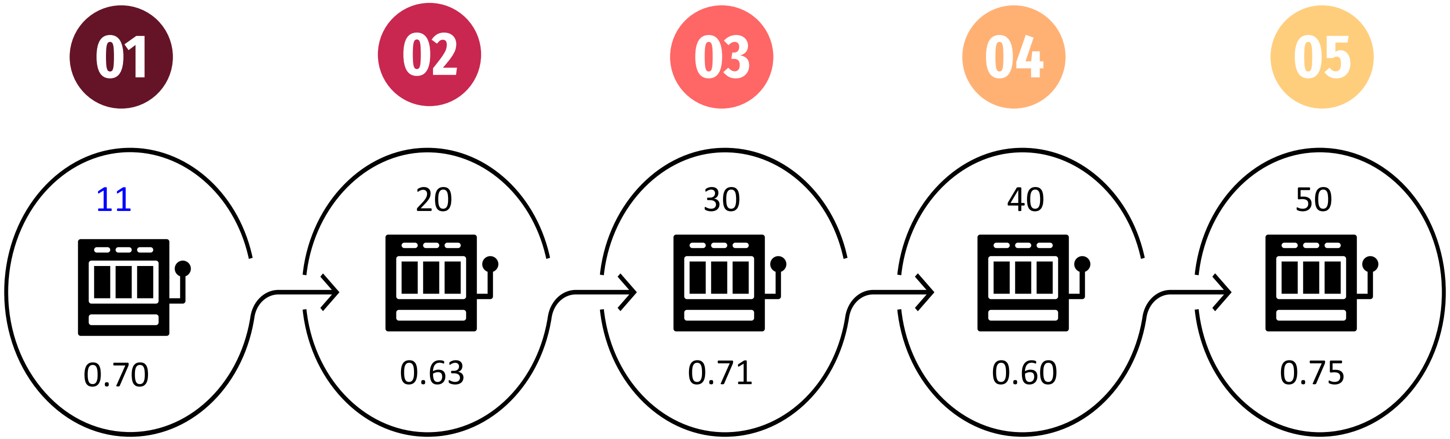
Best-Arm Identification Algorithm



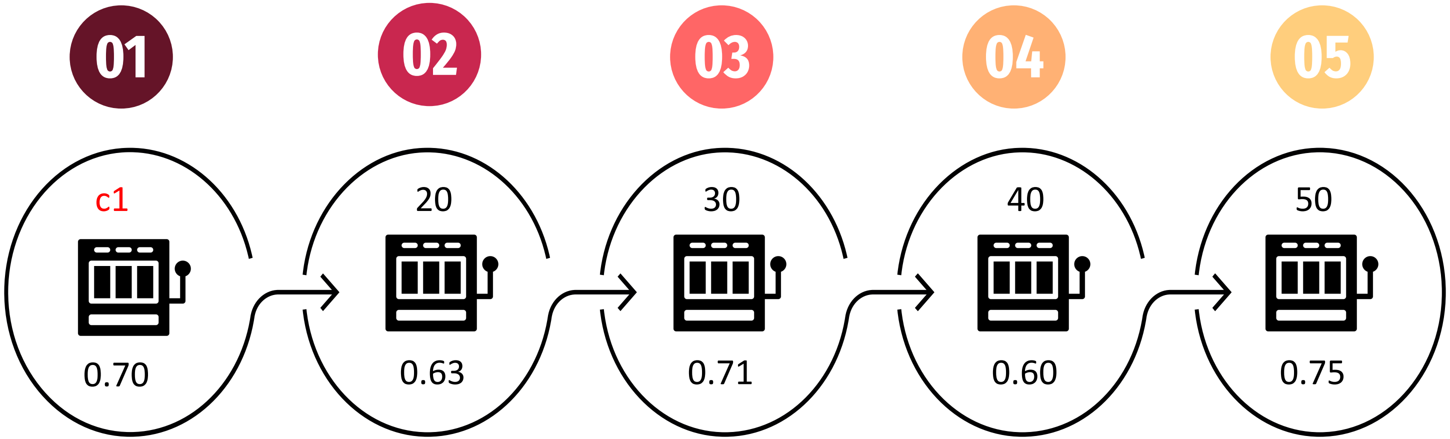
Best-Arm Identification Algorithm



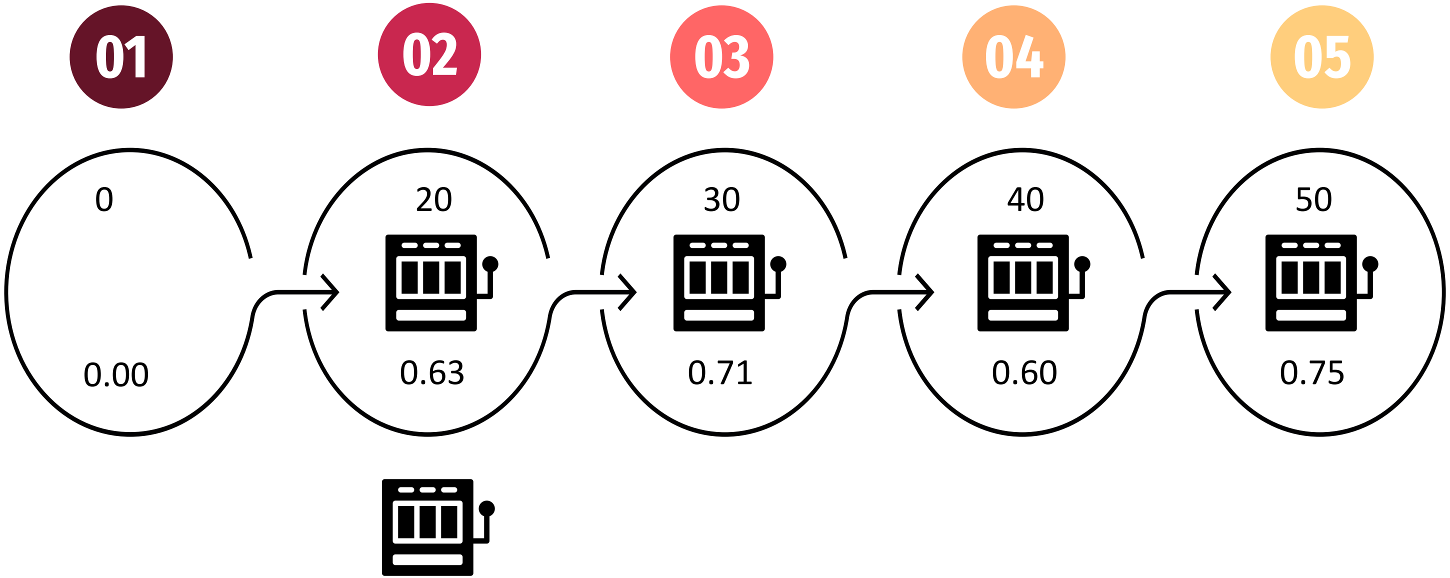
Best-Arm Identification Algorithm



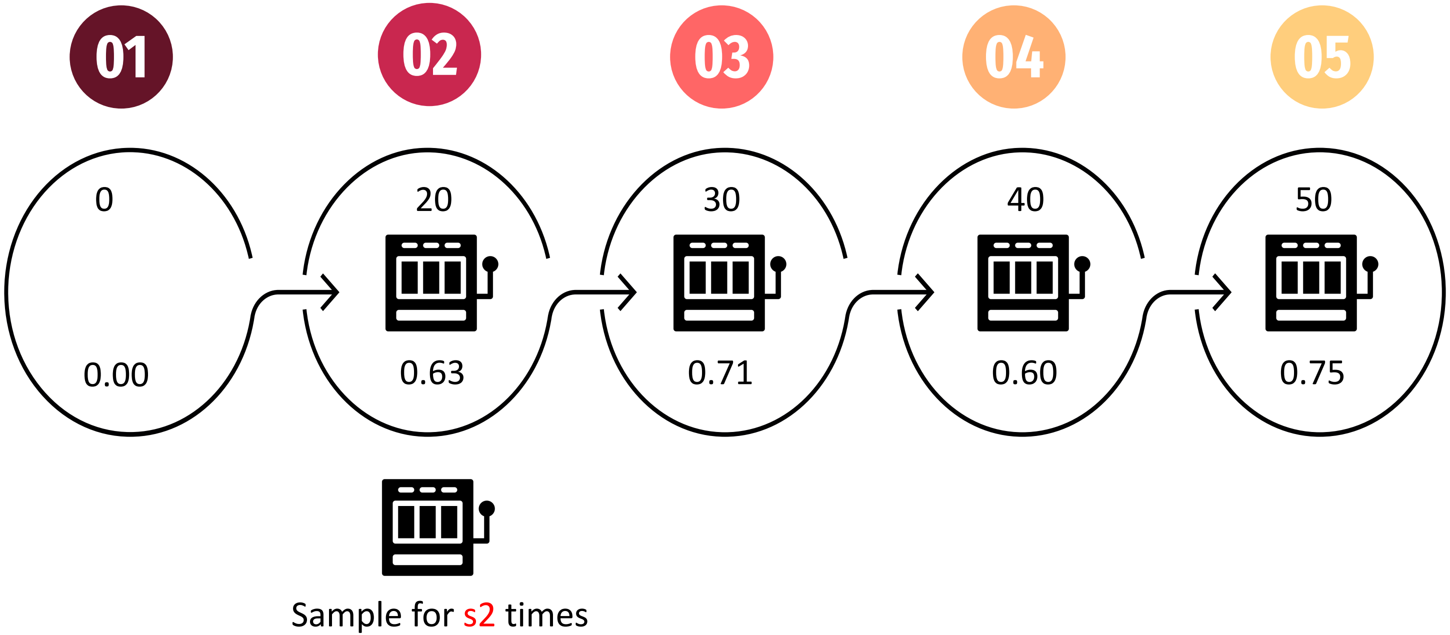
Best-Arm Identification Algorithm



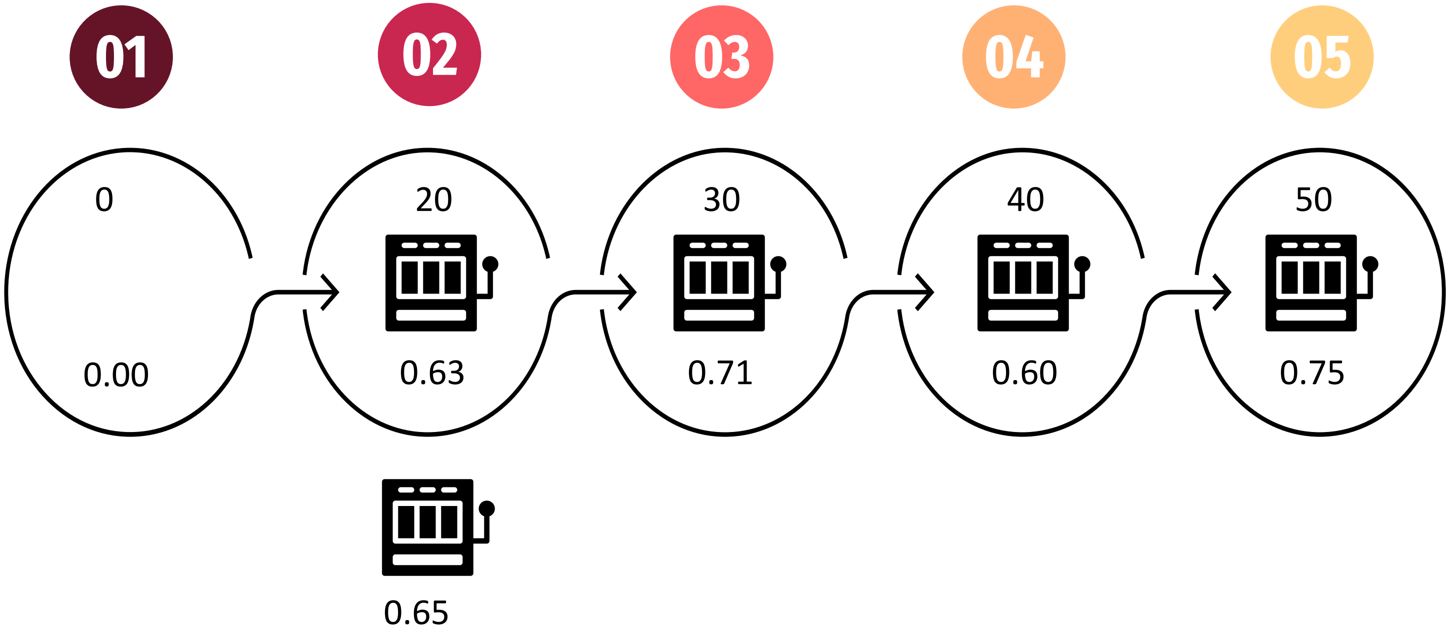
Best-Arm Identification Algorithm



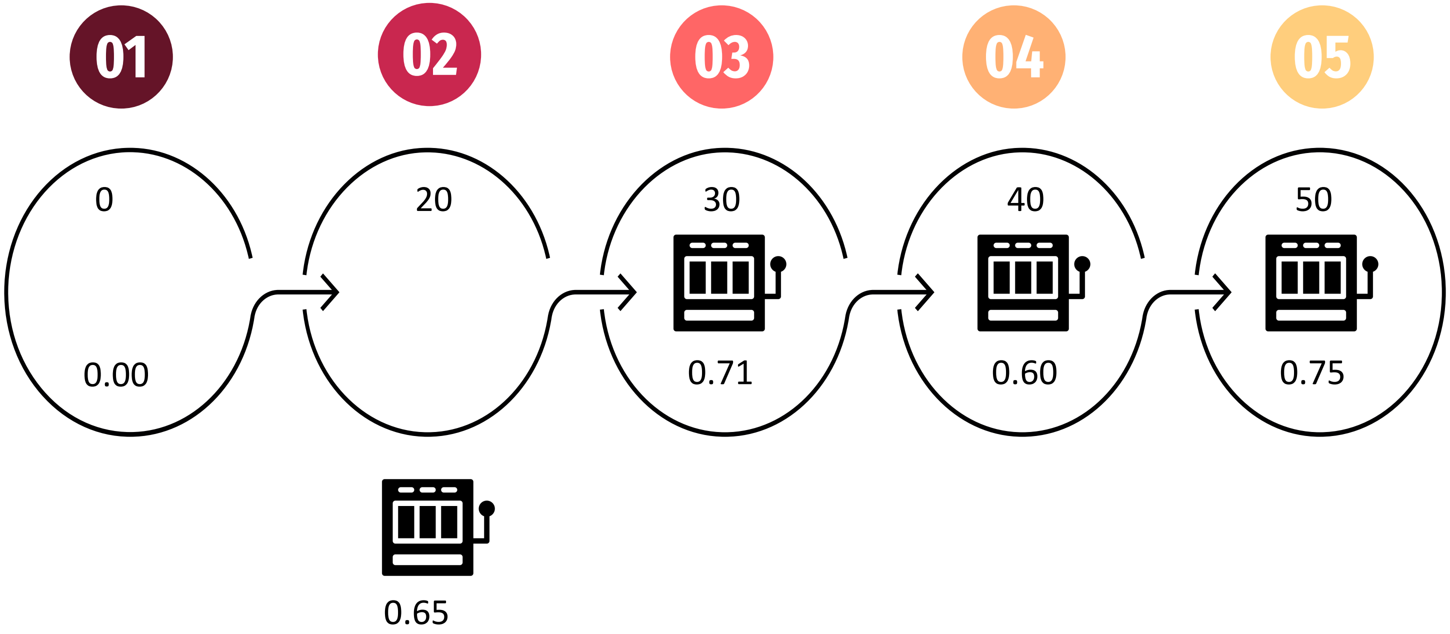
Best-Arm Identification Algorithm



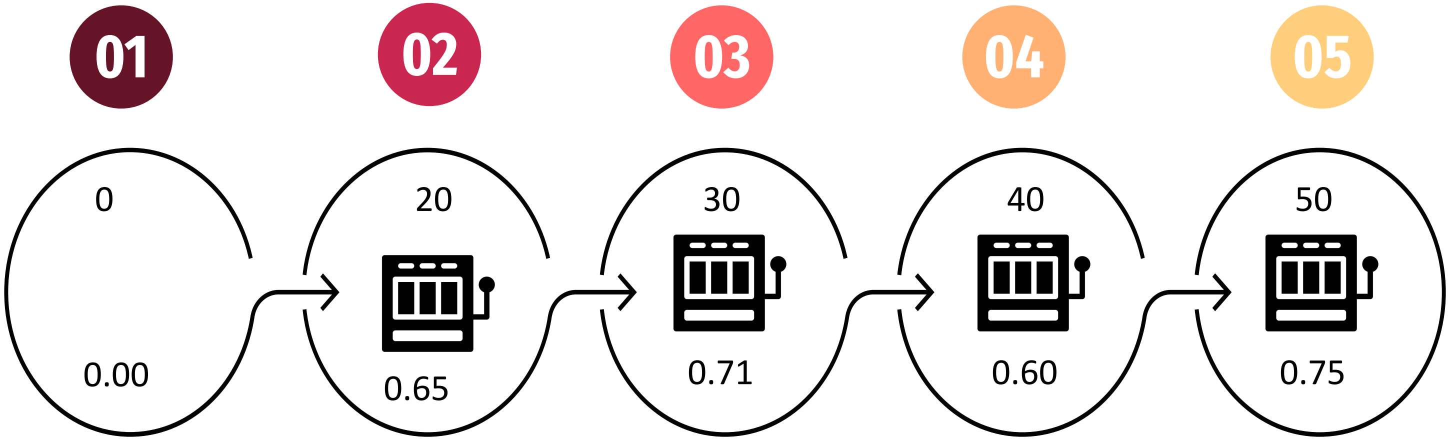
Best-Arm Identification Algorithm



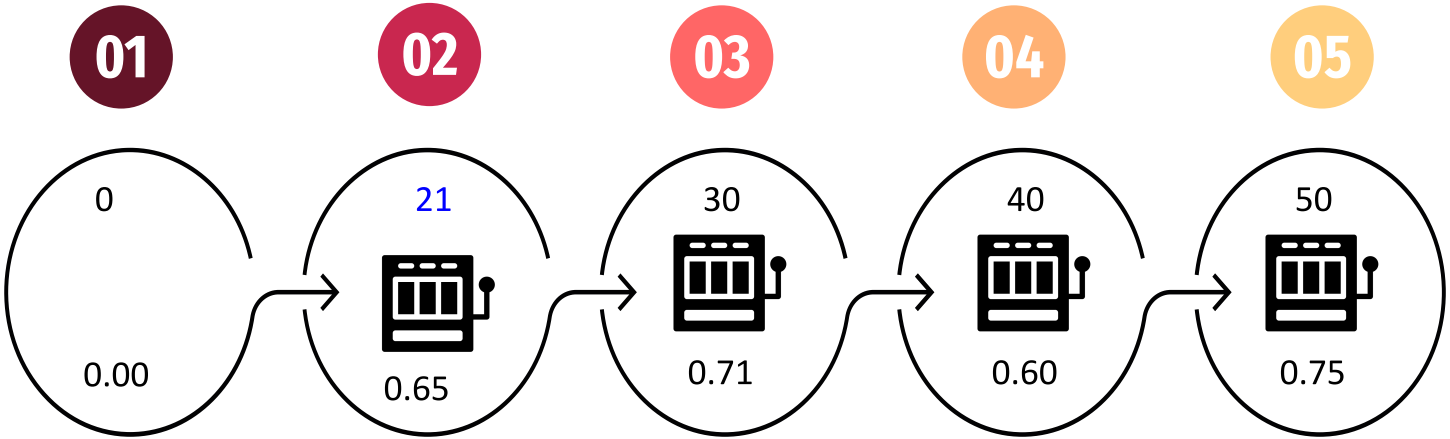
Best-Arm Identification Algorithm



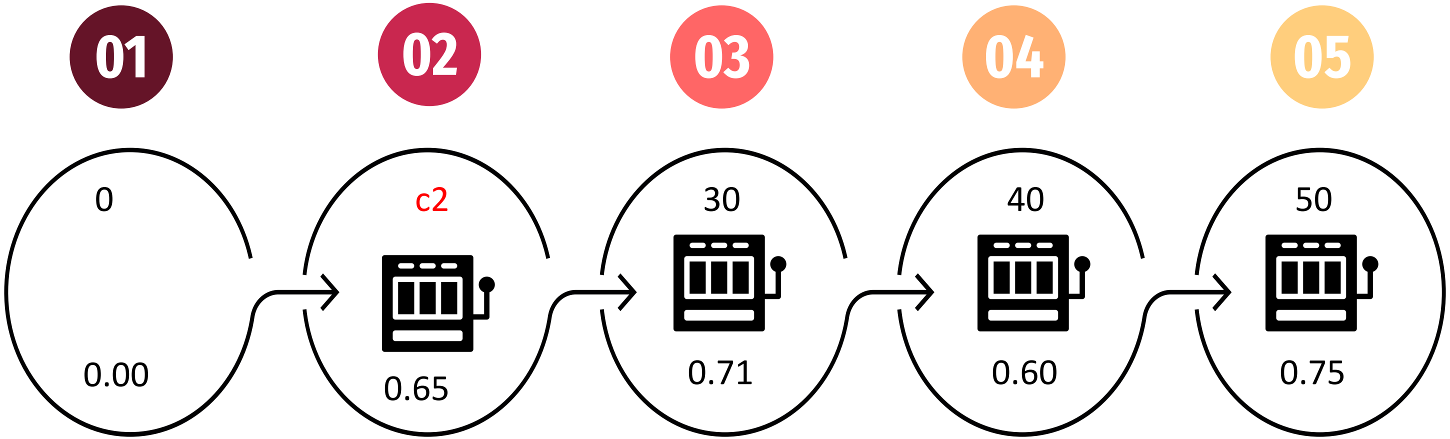
Best-Arm Identification Algorithm



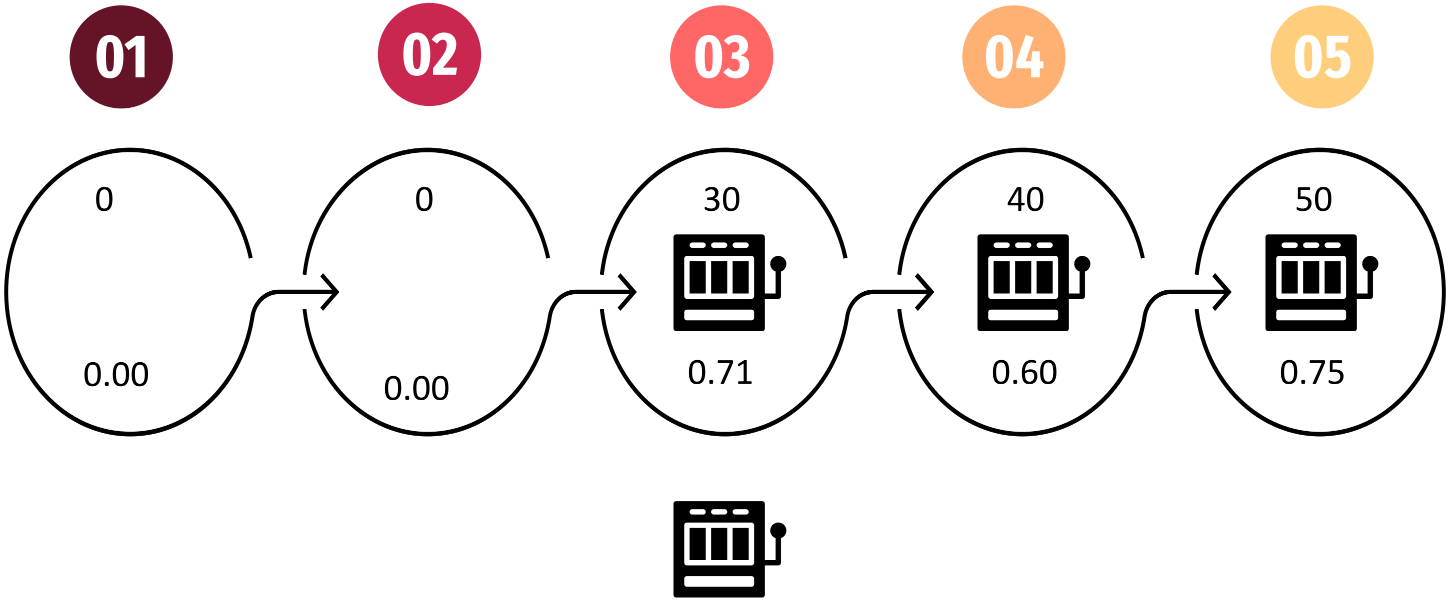
Best-Arm Identification Algorithm



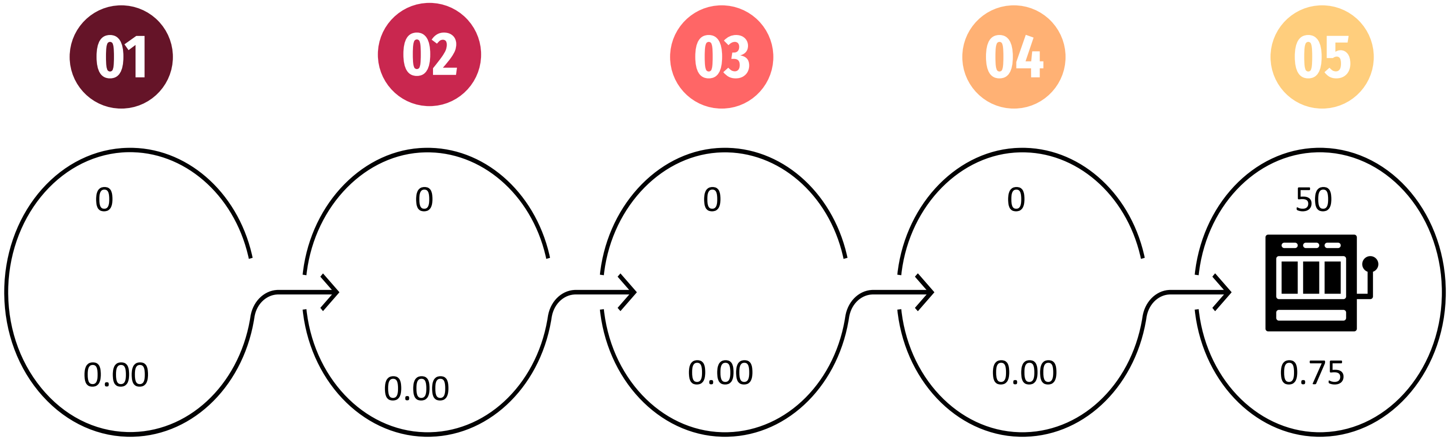
Best-Arm Identification Algorithm



Best-Arm Identification Algorithm



Best-Arm Identification Algorithm



Summary

Main Result 1: Any single-pass algorithm will incur at least

$$\Omega\left(\frac{n^{1/3}T^{2/3}}{m^{7/3}}\right)$$

expected cumulative regret when $m < n$.
This lower bound even holds for random order arrival.

$$O\left(\frac{n}{\varepsilon^2}\left(\mathtt{ilog}^{(\mathtt{r})}(n) + \log\left(\frac{1}{\delta}\right)\right)\right)$$

Main Result 2: (ε, δ) - PAC algorithm with

$O(r)$ arm memory

Optimal r-round sample complexity

Corollary: (ε, δ)- PAC algorithm with $O(\log^* n)$ arm memory with optimal worst-case sample complexity

Open Problems

1) Obtain instance-dependent lower bounds and upper bounds on the expected cumulative regret for single-pass MAB algorithms with bounded arm-memory.

2) Obtain lower bounds and upper bounds on the expected cumulative regret for k-pass MAB algorithms with bounded arm-memory where $k>1$.

3) Obtain an (ε, δ)-PAC streaming algorithm with $O(1)$ arm memory and optimal worst case sample complexity.

Open Problems

1) Obtain instance-dependent lower bounds and upper bounds on the expected cumulative regret for single-pass MAB algorithms with bounded arm-memory.

2) Obtain lower bounds and upper bounds on the expected cumulative regret for k-pass MAB algorithms with bounded arm-memory where $k>1$.

3) Obtain an (ε, δ)-PAC streaming algorithm with $O(1)$ arm memory and optimal worst case sample complexity.

Thank You