Cost Aware Best Arm Identification

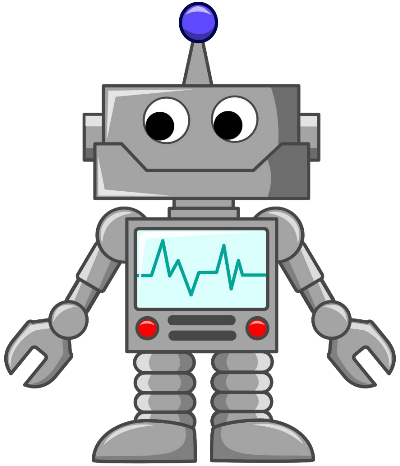
Reinforcement Learning Conference 2024

Kellen Kanarios, Qining Zhang, Lei Ying
{kellenkk, qiningz, leiying}@umich.edu

Outline

1. Recap MAB Motivations
2. Introduce New Formulation
3. Tight Asymptotic Lower Bound
4. Simple Algorithm
5. Results

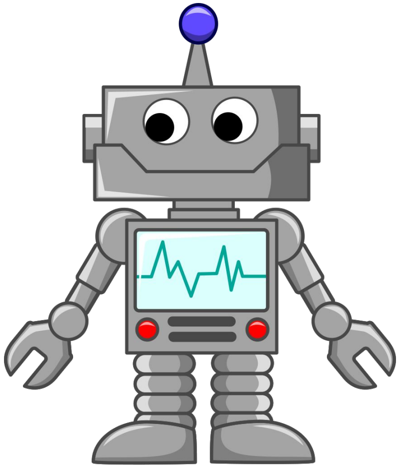
Scenario 1: Gambling Robot

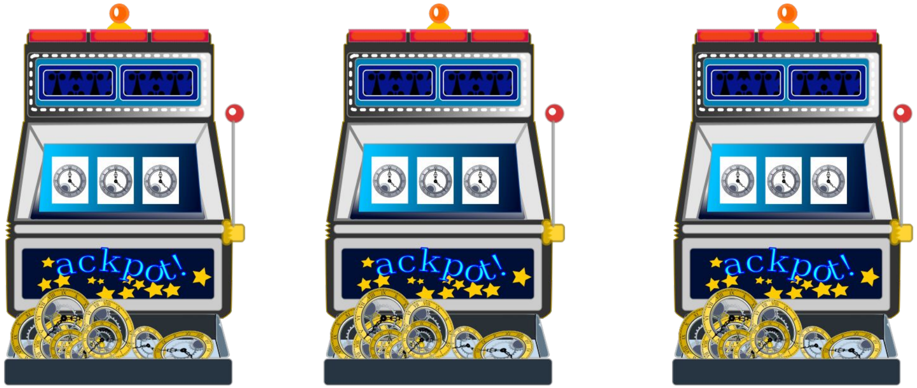


Scenario 1: Gambling Robot

$X_{i,t}$ = reward arm i at time t

$\mu_i = \mathbb{E}\left[X_{i,t}\right]$



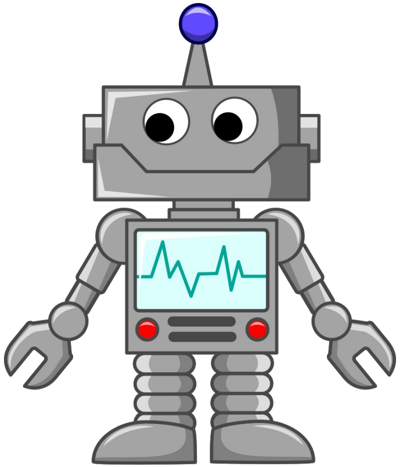


μ_1 μ_2 μ_3

Scenario 1: Gambling Robot

$X_{i,t}$ = reward arm i at time t

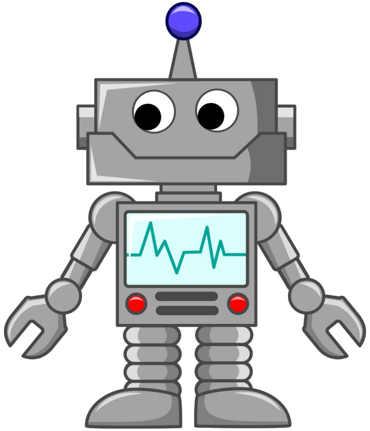
$\mu_i = \mathbb{E}\left[X_{i,t}\right]$



μ_1 μ_2 μ_3

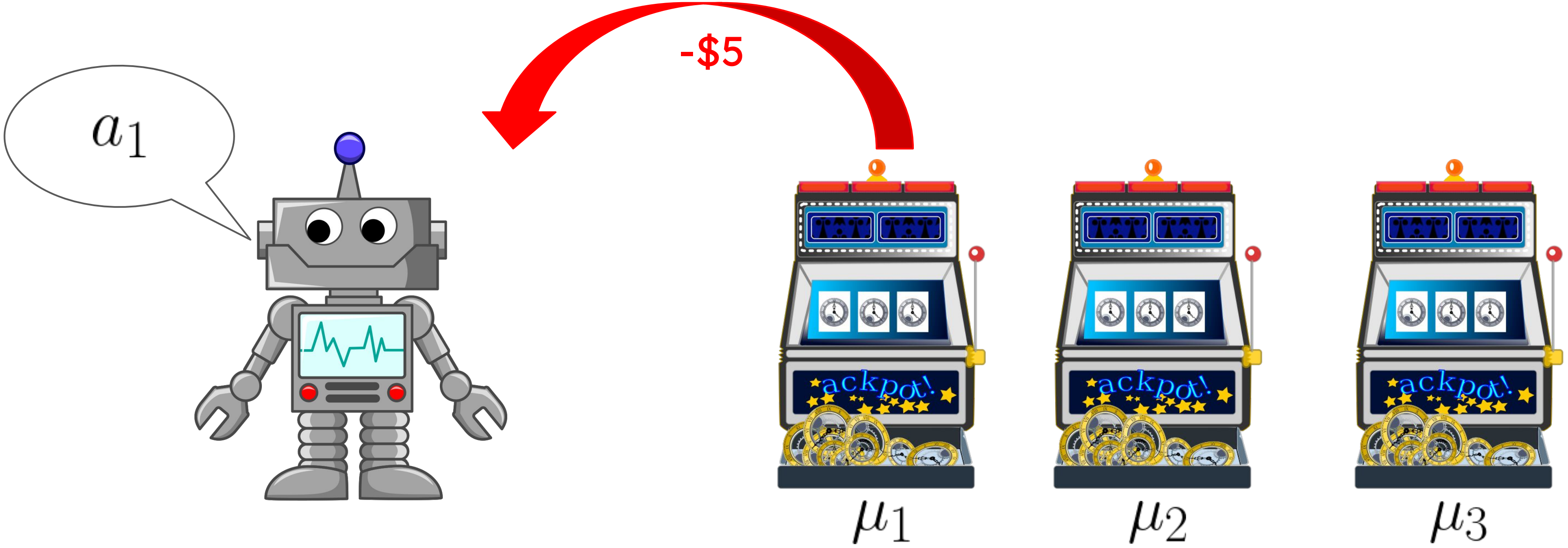
Goal: Make as much money as possible

Scenario 1: Gambling Robot

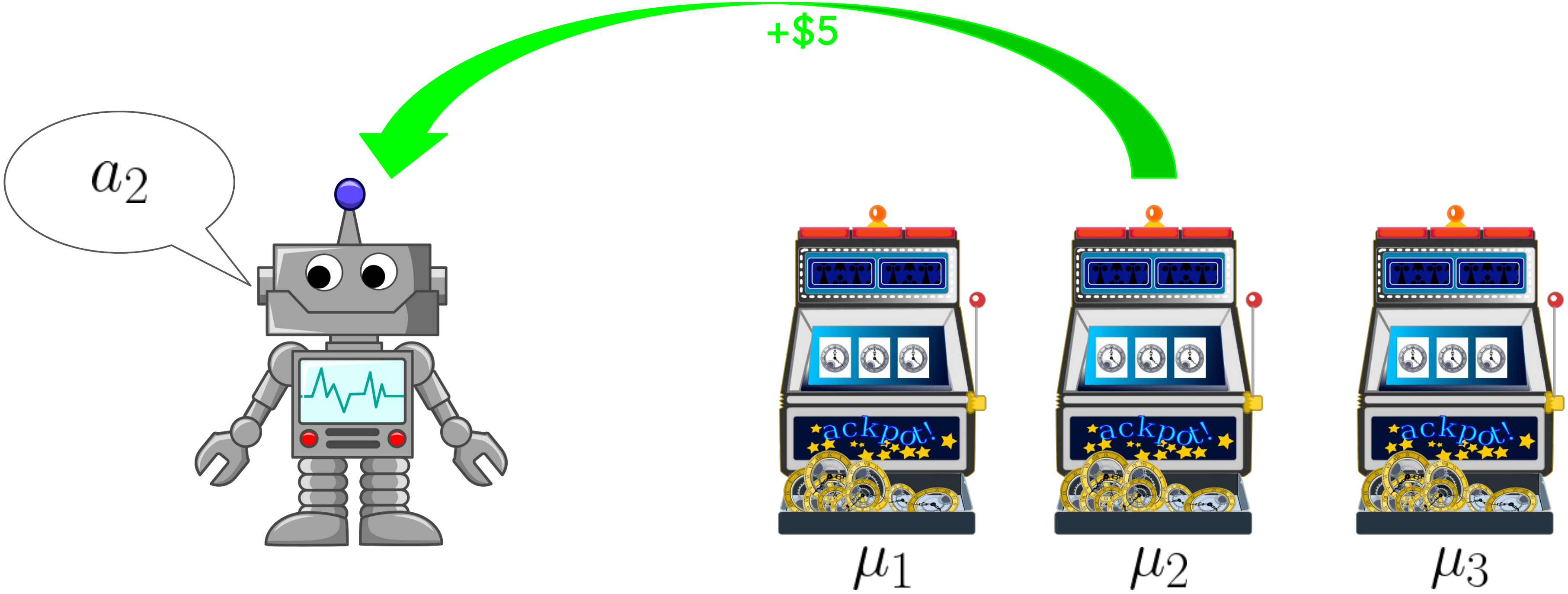


μ_1 μ_2 μ_3

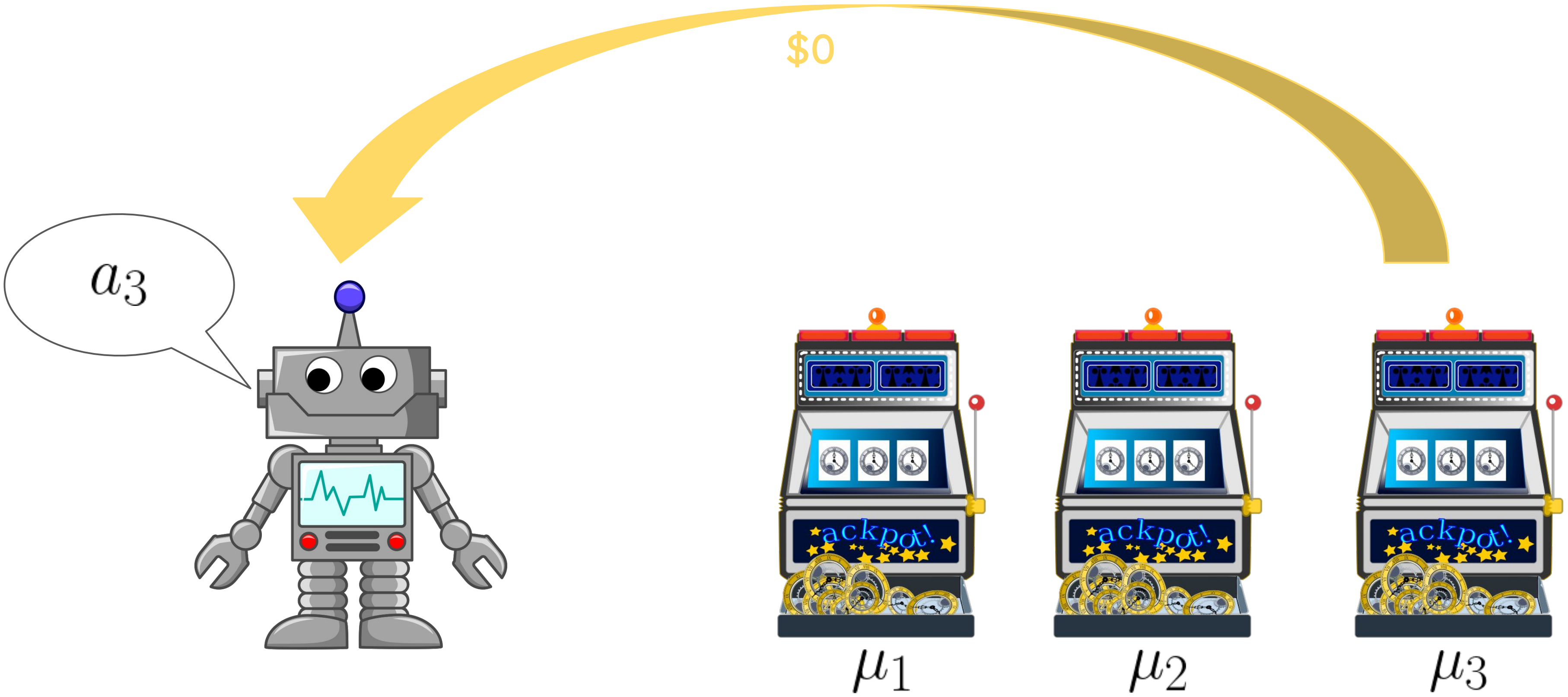
Scenario 1: Gambling Robot



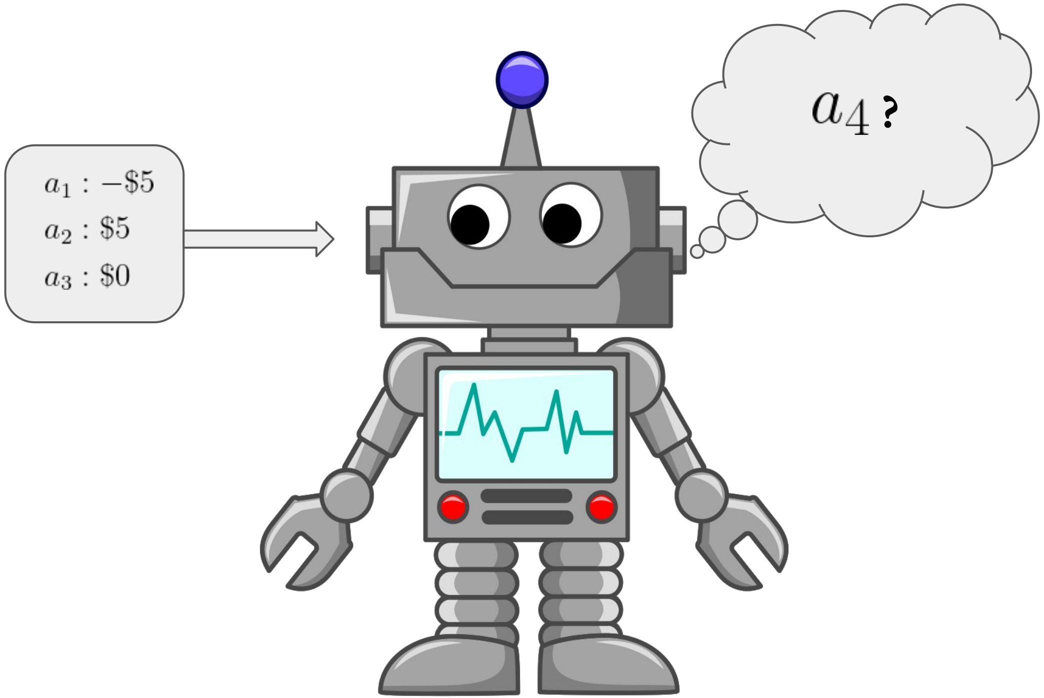
Scenario 1: Gambling Robot



Scenario 1: Gambling Robot



Scenario 1: Gambling Robot



Regret Minimization:

Inputs

Time Horizon

$$T \in \mathcal{N}$$

Components

Arm sampling rule

$$\pi : \mathcal{H} \to \mathcal{A}$$

Objective

Regret

$$\mathrm{Reg}_\pi(T) = \max_j \sum_{t=1}^{T} \mathbb{E}\left[X_{j,t} - X_{\pi(t),t}\right]$$

Scenario 2: Clinical Trials

Scenario 2: Clinical Trials

$S_i = i$th drug trial is a success.

$\mu_i = \mathbb{E}\left[\mathbf{1}_{\{S_i\}}\right]$

μ_1 μ_2 μ_3

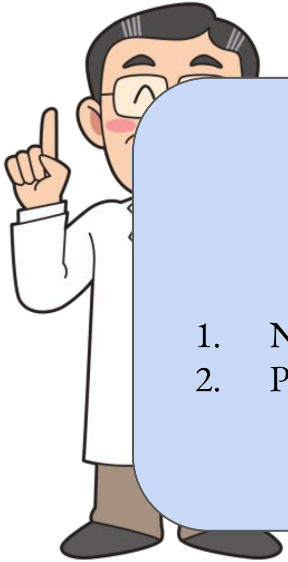
Scenario 2: Clinical Trials

$S_i = i$th drug trial is a success.
$\mu_i = \mathbb{E}\left[\mathbf{1}_{\{S_i\}}\right]$

μ_1 μ_2 μ_3

Goal: Identify best drug as fast as possible.

Scenario 2: Clinical Trials



Problem!

1. No fixed time horizon T.
2. Patients do not want to be the exploration arm pull.

μ_3

Best Arm Identification

Inputs

Time Horizon

$$T \in \mathcal{N}$$

Components

Arm sampling rule

$$\pi : \mathcal{H} \to \mathcal{A}$$

Objective

Regret

$$\mathrm{Reg}_\pi(T) = \max_j \sum_{t=1}^{T} \mathbb{E}\left[X_{j,t} - X_{\pi(t),t}\right]$$

Best Arm Identification

Inputs

Confidence Level

$$\delta \in [0, 1)$$

Components

Arm sampling rule

$$\pi : \mathcal{H} \to \mathcal{A}$$

Objective

Regret

$$\mathrm{Reg}_\pi(T) = \max_j \sum_{t=1}^{T} \mathbb{E}\left[X_{j,t} - X_{\pi(t),t}\right]$$

Best Arm Identification

Inputs

Confidence Level

$$\delta \in [0, 1)$$

Components

Arm sampling rule

$$\pi : \mathcal{H} \to \mathcal{A}$$

Stopping Rule

$$\tau : \mathcal{H} \to \{0, 1\}$$

Objective

Regret

$$\mathrm{Reg}_\pi(T) = \max_j \sum_{t=1}^{T} \mathbb{E}\left[X_{j,t} - X_{\pi(t),t}\right]$$

Best Arm Identification

Inputs

Confidence Level

$$\delta \in [0, 1)$$

Components

Arm sampling rule

$$\pi : \mathcal{H} \to \mathcal{A}$$

Stopping Rule

$$\tau : \mathcal{H} \to \{0, 1\}$$

Objective

Sample Complexity

$$\min \tau$$
$$\text{s.t } \Pr(\hat{a} = a^*) \geq 1 - \delta$$

Scenario 3: Corporate Health

Scenario 3: Corporate Health

$$c_i = \mathbb{E}\left[\text{cost of trial for drug } i\right]$$

c_1

μ_1

c_2

μ_2

c_3

μ_3

Scenario 3: Corporate Health

$$c_i = \mathbb{E}\left[\text{cost of trial for drug } i\right]$$

c_1 c_2 c_3

μ_1 μ_2 μ_3

Goal: Identify best drug as cheap as possible.

Cost Aware Best Arm Identification

Inputs

Confidence Level

$$\delta \in [0, 1)$$

Components

Arm sampling rule

$$\pi : \mathcal{H} \to \mathcal{A}$$

Stopping Rule

$$\tau : \mathcal{H} \to \{0, 1\}$$

Objective

Sample Complexity

$$\min \tau$$
$$\text{s.t } \Pr(\hat{a} = a^*) \geq 1 - \delta$$

Kanarios, Kellen; Zhang, Qining; Ying, Lei. (2024). Cost Aware Best Arm Identification.

Cost Aware Best Arm Identification

Inputs

Confidence Level

$$\delta \in [0, 1)$$

Components

Arm sampling rule

$$\pi : \mathcal{H} \to \mathcal{A}$$

Stopping Rule

$$\tau : \mathcal{H} \to \{0, 1\}$$

Objective (Ours)

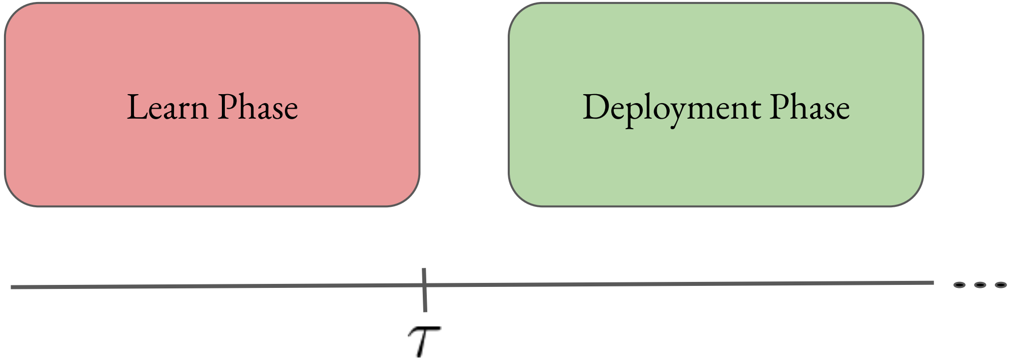
Cost Complexity

$$\min \sum_a \mathbb{E}\left[c_a N_a(\tau_\delta)\right]$$

$$\text{s.t } \Pr(\hat{a} = a^*) \geq 1 - \delta$$

Kanarios, Kellen; Zhang, Qining; Ying, Lei. (2024). Cost Aware Best Arm Identification.

Cost Aware Best Arm Identification



Objective (Ours)

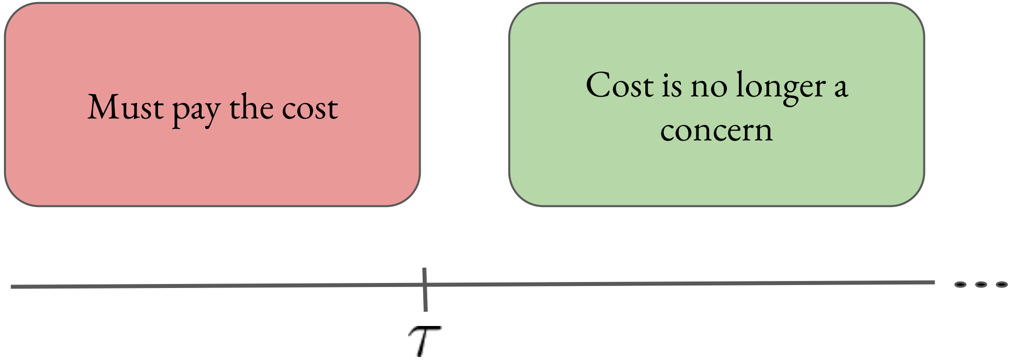
Cost Complexity

$$\min \sum_a \mathbb{E}\left[c_a N_a(\tau_\delta)\right]$$

$$\text{s.t } \Pr(\hat{a} = a^*) \geq 1 - \delta$$

Kanarios, Kellen; Zhang, Qining; Ying, Lei. (2024). Cost Aware Best Arm Identification.

Cost Aware Best Arm Identification



Objective (Ours)

Cost Complexity

$$\min \sum_a \mathbb{E}\left[c_a N_a(\tau_\delta)\right]$$

$$\text{s.t } \Pr(\hat{a} = a^*) \geq 1 - \delta$$

Kanarios, Kellen; Zhang, Qining; Ying, Lei. (2024). Cost Aware Best Arm Identification.

Cost Aware Best Arm Identification

Must pay the cost:
Clinical Trial

Cost is no longer a concern:
Cost is covered by insurance

τ ...

Objective (Ours)

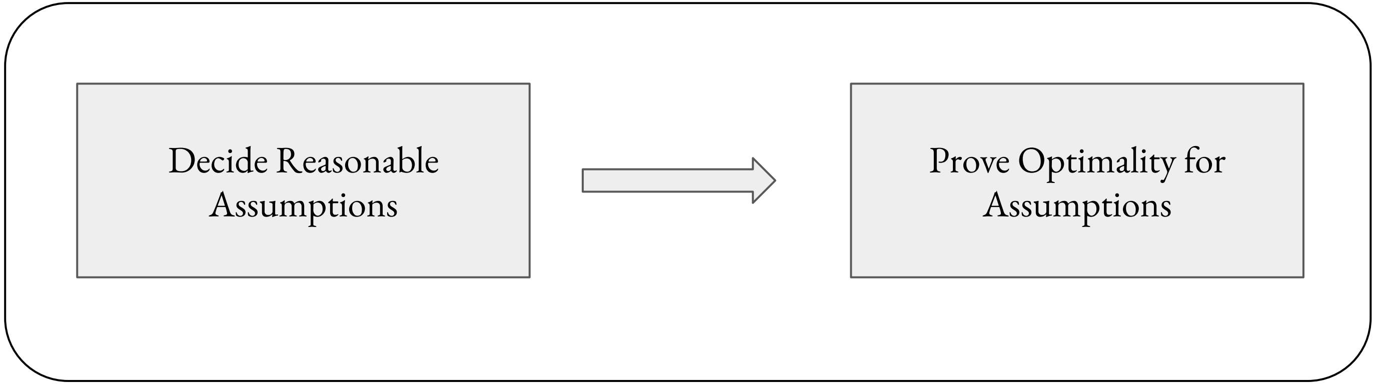
Cost Complexity

$$\min \sum_a \mathbb{E}\left[c_a N_a(\tau_\delta)\right]$$

$$\text{s.t } \Pr(\hat{a} = a^*) \geq 1 - \delta$$

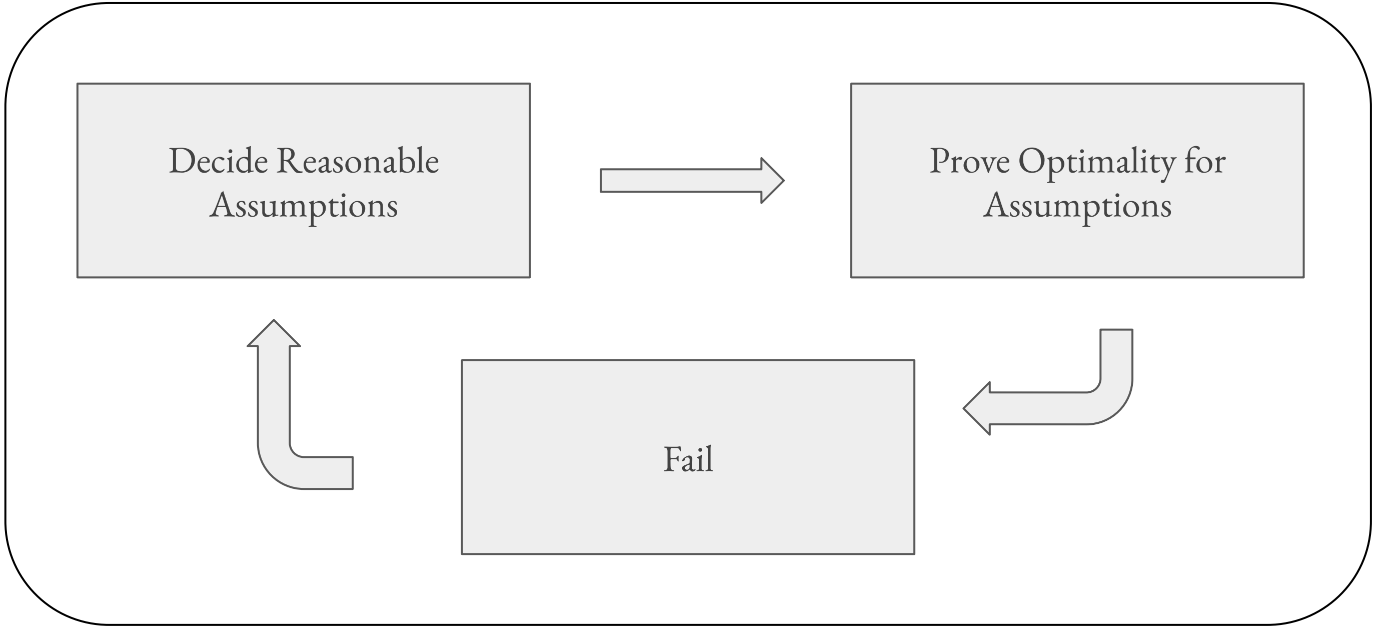
Kanarios, Kellen; Zhang, Qining; Ying, Lei. (2024). Cost Aware Best Arm Identification.

Optimal Cost Aware Best Arm Identification



Kanarios, Kellen; Zhang, Qining; Ying, Lei. (2024). Cost Aware Best Arm Identification.

Optimal Cost Aware Best Arm Identification



Kanarios, Kellen; Zhang, Qining; Ying, Lei. (2024). Cost Aware Best Arm Identification.

Optimal Cost Aware Best Arm Identification

Assumptions on Rewards

1. Single parameter exponential family.
2. Unique best arm.

Inherited from previous BAI work.

Kanarios, Kellen; Zhang, Qining; Ying, Lei. (2024). Cost Aware Best Arm Identification.

Optimal Cost Aware Best Arm Identification

Assumptions on Rewards

1. Single parameter exponential family.
2. Unique best arm.

Assumptions on Costs

1. Costs are bounded.
2. Costs are strictly greater than 0.

Inherited from previous BAI work.

Ours

Kanarios, Kellen; Zhang, Qining; Ying, Lei. (2024). Cost Aware Best Arm Identification.

Optimal Cost Aware Best Arm Identification

"Everything has a price"

"Nothing in life is free"

Assumptions on Costs

1. Costs are bounded.
2. Costs are strictly greater than 0.

Ours

Kanarios, Kellen; Zhang, Qining; Ying, Lei. (2024). Cost Aware Best Arm Identification.

Optimal Cost Aware Best Arm Identification

Lower Bound. Let $\delta \in (0,1)$. For any δ - PAC algorithm and any bandit model $\boldsymbol{\mu} \in \mathcal{M}$, we have:

$$\sum_a \mathbb{E}_{\boldsymbol{\mu} \times \boldsymbol{c}}\left[c_a N_a\left(\tau_\delta\right)\right] \geq T^*(\boldsymbol{\mu}) \log \frac{1}{\delta} + o\left(\log \frac{1}{\delta}\right)$$

where $T^*(\boldsymbol{\mu})$ is the instance dependent constant satisfying:

$$T^*(\boldsymbol{\mu})^{-1} = \sup_{\boldsymbol{w} \in \Sigma_K} \inf_{\boldsymbol{\lambda} \in \{a^*(\boldsymbol{\lambda}) \neq a^*(\boldsymbol{\mu})\}} \sum_a \frac{w_a}{c_a} d\left(\mu_a, \lambda_a\right).$$

Kanarios, Kellen; Zhang, Qining; Ying, Lei. (2024). Cost Aware Best Arm Identification.

Stopping Rule

In CABAI, we still want to stop as soon as possible

- Can reuse **exact** stopping rule from BAI!

Garivier, A; Kaufmann, E. (2016). Optimal Best Arm Identification with Fixed Confidence.

Stopping Rule

In CABAI, we still want to stop as soon as possible

- Can reuse **exact** stopping rule from BAI!

$$\hat{\mu}_{a,b}(t) := \frac{N_a(t)}{N_a(t) + N_b(t)}\hat{\mu}_a(t) + \frac{N_b(t)}{N_a(t) + N_b(t)}\hat{\mu}_b(t)$$

$$Z_{a,b}(t) = N_a(t)d\left(\hat{\mu}_a(t), \hat{\mu}_{a,b}(t)\right) + N_b(t)d\left(\hat{\mu}_b(t), \hat{\mu}_{a,b}(t)\right)$$

$$\tau_\delta = \inf\left\{t \in \mathbb{N} : Z(t) := \max_{a\in\mathcal{A}} \min_{b\in\mathcal{A}\setminus\{a\}} Z_{a,b}(t) > \beta(t, \delta)\right\}$$

Garivier, A; Kaufmann, E. (2016). Optimal Best Arm Identification with Fixed Confidence.

Sampling Rule

Proof Snippet:

$$\begin{aligned}\mathrm{kl}(\delta, 1-\delta) &\leq \mathbb{E}_{\boldsymbol{\mu}\times\boldsymbol{c}}\left[C\left(\tau_\delta\right)\right] \inf_{\lambda\in\mathrm{Alt}(\boldsymbol{\mu})}\left(\sum_{a=1}^{K}\frac{\mathbb{E}_{\boldsymbol{\mu}\times\boldsymbol{c}}\left[c_a N_a\left(\tau_\delta\right)\right]}{\mathbb{E}_{\boldsymbol{\mu}\times\boldsymbol{c}}\left[c_a C\left(\tau_\delta\right)\right]} d\left(\mu_a, \lambda_a\right)\right)\\ &\leq \mathbb{E}_{\boldsymbol{\mu}}\left[C\left(\tau_\delta\right)\right] \sup_{w\in\Sigma_K}\inf_{\lambda\in\mathrm{Alt}(\boldsymbol{\mu})}\left(\sum_{a=1}^{K}\frac{w_a}{c_a} d\left(\mu_a, \lambda_a\right)\right),\end{aligned}$$

Kanarios, Kellen; Zhang, Qining; Ying, Lei. (2024). Cost Aware Best Arm Identification.

Sampling Rule

Proof Snippet:

$$\mathrm{kl}(\delta, 1-\delta) \leq \mathbb{E}_{\boldsymbol{\mu}\times\boldsymbol{c}}\left[C\left(\tau_\delta\right)\right] \inf_{\lambda\in\mathrm{Alt}(\boldsymbol{\mu})}\left(\sum_{a=1}^{K}\frac{\mathbb{E}_{\boldsymbol{\mu}\times\boldsymbol{c}}\left[c_a N_a\left(\tau_\delta\right)\right]}{\mathbb{E}_{\boldsymbol{\mu}\times\boldsymbol{c}}\left[c_a C\left(\tau_\delta\right)\right]} d\left(\mu_a, \lambda_a\right)\right)$$

$$\leq \mathbb{E}_{\boldsymbol{\mu}}\left[C\left(\tau_\delta\right)\right] \boxed{\sup_{w\in\Sigma_K}} \inf_{\lambda\in\mathrm{Alt}(\boldsymbol{\mu})}\left(\sum_{a=1}^{K}\frac{w_a}{c_a} d\left(\mu_a, \lambda_a\right)\right),$$

Key insight 1: Can compute $\boxed{w^*(\boldsymbol{\mu})}$ for given $\boldsymbol{\mu}$

Kanarios, Kellen; Zhang, Qining; Ying, Lei. (2024). Cost Aware Best Arm Identification.

Sampling Rule

Proof Snippet:

$$\begin{aligned}\mathrm{kl}(\delta, 1-\delta) &\leq \mathbb{E}_{\boldsymbol{\mu}\times\boldsymbol{c}}\left[C\left(\tau_\delta\right)\right] \inf_{\lambda\in\mathrm{Alt}(\boldsymbol{\mu})}\left(\sum_{a=1}^{K}\boxed{\frac{\mathbb{E}_{\boldsymbol{\mu}\times\boldsymbol{c}}\left[c_a N_a\left(\tau_\delta\right)\right]}{\mathbb{E}_{\boldsymbol{\mu}\times\boldsymbol{c}}\left[c_a C\left(\tau_\delta\right)\right]}} d\left(\mu_a, \lambda_a\right)\right) \\ &\leq \mathbb{E}_{\boldsymbol{\mu}}\left[C\left(\tau_\delta\right)\right] \boxed{\sup_{w\in\Sigma_K}} \inf_{\lambda\in\mathrm{Alt}(\boldsymbol{\mu})}\left(\sum_{a=1}^{K}\frac{w_a}{c_a} d\left(\mu_a, \lambda_a\right)\right),\end{aligned}$$

Key insight 1: Can compute $\boxed{w^*(\boldsymbol{\mu})}$ for given $\boldsymbol{\mu}$

Key insight 2: Proportion should match $w^*(\boldsymbol{\mu})$

Kanarios, Kellen; Zhang, Qining; Ying, Lei. (2024). Cost Aware Best Arm Identification.

Sampling Rule

Proof Snippet:

$$\mathrm{kl}(\delta, 1-\delta) \leq \mathbb{E}_{\boldsymbol{\mu}\times\boldsymbol{c}}\left[C\left(\tau_\delta\right)\right] \inf_{\lambda \in \mathrm{Alt}(\boldsymbol{\mu})}\left(\sum_{a=1}^{K} \boxed{\frac{\mathbb{E}_{\boldsymbol{\mu}\times\boldsymbol{c}}\left[c_a N_a\left(\tau_\delta\right)\right]}{\mathbb{E}_{\boldsymbol{\mu}\times\boldsymbol{c}}\left[c_a C\left(\tau_\delta\right)\right]}} d\left(\mu_a, \lambda_a\right)\right)$$

$$\leq \mathbb{E}_{\boldsymbol{\mu}}\left[C\left(\tau_\delta\right)\right] \boxed{\sup_{w \in \Sigma_K}} \inf_{\lambda \in \mathrm{Alt}(\boldsymbol{\mu})}\left(\sum_{a=1}^{K} \frac{w_a}{c_a} d\left(\mu_a, \lambda_a\right)\right),$$

Key insight 1: Can compute $\boxed{w^*(\boldsymbol{\mu})}$ for given $\boldsymbol{\mu}$

Key insight 2: Proportion should match $w^*(\boldsymbol{\mu})$

Sampling Rule: $\pi(t) \in \arg\max_i \left|C(t) w^*(\widehat{\boldsymbol{\mu}}(t)) - \widehat{c}_i(t) N_i(t)\right|$

Kanarios, Kellen; Zhang, Qining; Ying, Lei. (2024). Cost Aware Best Arm Identification.

Optimal Cost Aware Best Arm Identification

$$\pi(t) \in \arg\max_i \left| C(t) w^*(\widehat{\boldsymbol{\mu}}(t)) - \widehat{c}_i(t) N_i(t) \right|$$

+

$$\tau_\delta = \inf \left\{ t \in \mathbb{N} : Z(t) := \max_{a \in \mathcal{A}} \min_{b \in \mathcal{A} \setminus \{a\}} Z_{a,b}(t) > \beta(t, \delta) \right\}$$

Kanarios, Kellen; Zhang, Qining; Ying, Lei. (2024). Cost Aware Best Arm Identification.

Optimal Cost Aware Best Arm Identification

$$\pi(t) \in \arg\max_i \left|C(t)w^*(\widehat{\boldsymbol{\mu}}(t)) - \widehat{c}_i(t)N_i(t)\right|$$

$+$

$$\tau_\delta = \inf\left\{t \in \mathbb{N} : Z(t) := \max_{a\in\mathcal{A}} \min_{b\in\mathcal{A}\setminus\{a\}} Z_{a,b}(t) > \beta(t,\delta)\right\}$$

Asymptotic Optimality in Expectation.

For suitably chosen $\beta(t,\delta)$ we have

$$\limsup_{\delta\to 0} \frac{\mathbb{E}\left[C(\tau_\delta)\right]}{\log(1/\delta)} \leq T^*(\boldsymbol{\mu})$$

Kanarios, Kellen; Zhang, Qining; Ying, Lei. (2024). Cost Aware Best Arm Identification.

Optimal Cost Aware Best Arm Identification

$$\pi(t) \in \arg\max_i \left| C(t) w^*(\widehat{\boldsymbol{\mu}}(t)) - \widehat{c_i}(t) N_i(t) \right|$$

$$\tau_\delta = \inf \left\{ t \in \mathbb{N} : Z(t) := \max_{a \in \mathcal{A}} \min_{b \in \mathcal{A} \setminus \{a\}} Z_{a,b}(t) > \beta(t, \delta) \right\}$$

Algorithm	Optimal?	$w_1(t)$ $(1.5, 1)$	$w_2(t)$ $(1, 0.1)$	$w_3(t)$ $(0.5, 0.01)$
TAS	$\times$	0.46	0.46	0.08
CTAS	✓	0.23	0.72	0.05

Optimal proportions of Cost Aware vs. non Cost Aware algorithm.

Kanarios, Kellen; Zhang, Qining; Ying, Lei. (2024). Cost Aware Best Arm Identification.

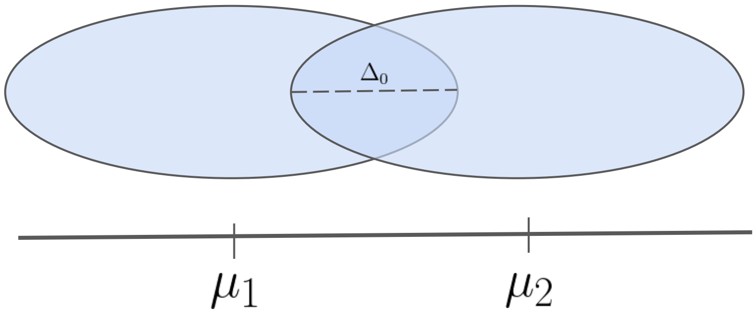
A heuristic: $\sqrt{c_i}$

- Solving lower bound optimization - HARD
- Computing confidence regions - EASY

Kanarios, Kellen; Zhang, Qining; Ying, Lei. (2024). Cost Aware Best Arm Identification.

A heuristic: $\sqrt{c_i}$

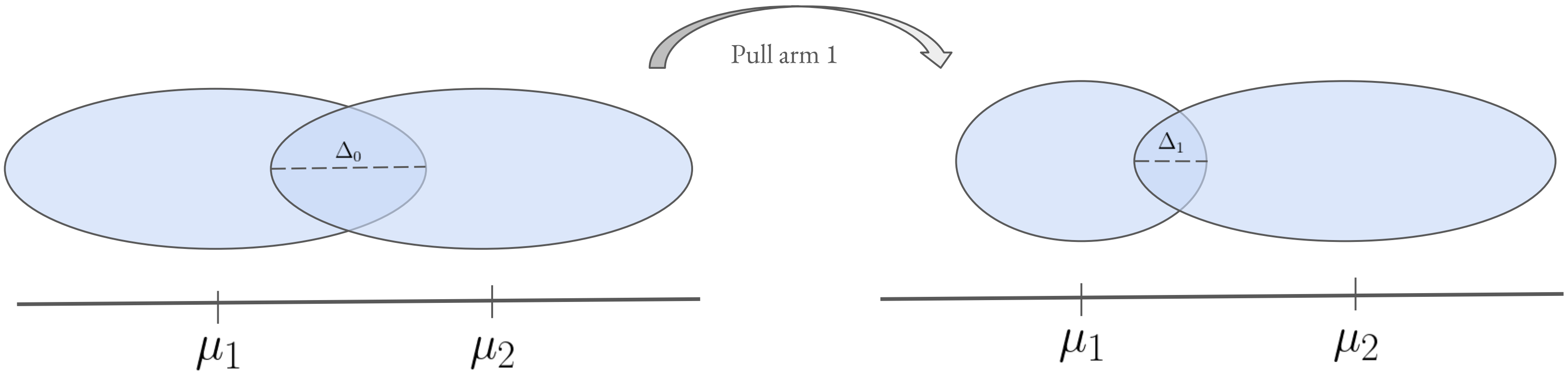
- Solving lower bound optimization - HARD
- Computing confidence regions - EASY



Kanarios, Kellen; Zhang, Qining; Ying, Lei. (2024). Cost Aware Best Arm Identification.

A heuristic: $\sqrt{c_i}$

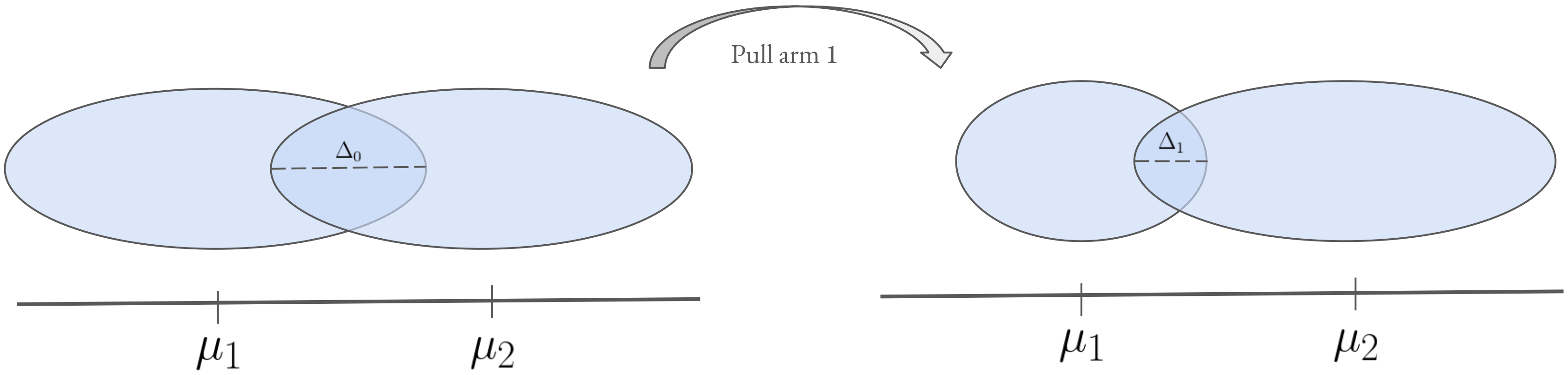
- Solving lower bound optimization - HARD
- Computing confidence regions - EASY



Kanarios, Kellen; Zhang, Qining; Ying, Lei. (2024). Cost Aware Best Arm Identification.

A heuristic: $\sqrt{c_i}$

- **Idea**: Most reduced overlap for the cost $\dfrac{\Delta_1 - \Delta_0}{c_1}$



Kanarios, Kellen; Zhang, Qining; Ying, Lei. (2024). Cost Aware Best Arm Identification.

A heuristic: $\sqrt{c_i}$

- **Idea**: Most reduced overlap for the cost $\dfrac{\Delta_1 - \Delta_0}{c_1}$

Sampling Rule:

$$a_t \in \arg\max_i \frac{\Delta_{t+1}^{\hat{a},a_i} - \Delta_t^{\hat{a},a_i}}{c_i}$$

Elimination Rule:

Eliminate i at t if $\Delta_t^{\hat{a},a_i} = 0$

Stopping Rule:

Stop at t if $\max_i \Delta_t^{\hat{a},a_i} = 0$

Kanarios, Kellen; Zhang, Qining; Ying, Lei. (2024). Cost Aware Best Arm Identification.

A heuristic: $\sqrt{c_i}$

- **Observation**: sampling rule can be approximated by $a_t \in \arg\max_i \sqrt{c_i} N_i(t)$

Theoretical Result. For any 2-armed Gaussian bandit model with rewards $\{\mu_1, \mu_2\}$ and costs $\{c_1, c_2\}$, we have with probability 1:

$$\limsup_{\delta \to 0} \frac{\sum_a \mathbb{E}\left[c_a N_a(\tau_\delta)\right]}{\log(1/\delta)} \leq \frac{2\alpha\left(\sqrt{c_1} + \sqrt{c_2}\right)^2}{(\mu_1 - \mu_2)^2}$$

Kanarios, Kellen; Zhang, Qining; Ying, Lei. (2024). Cost Aware Best Arm Identification.

A heuristic: $\sqrt{c_i}$

- **Observation**: sampling rule can be approximated by $a_t \in \arg\max_i \sqrt{c_i} N_i(t)$

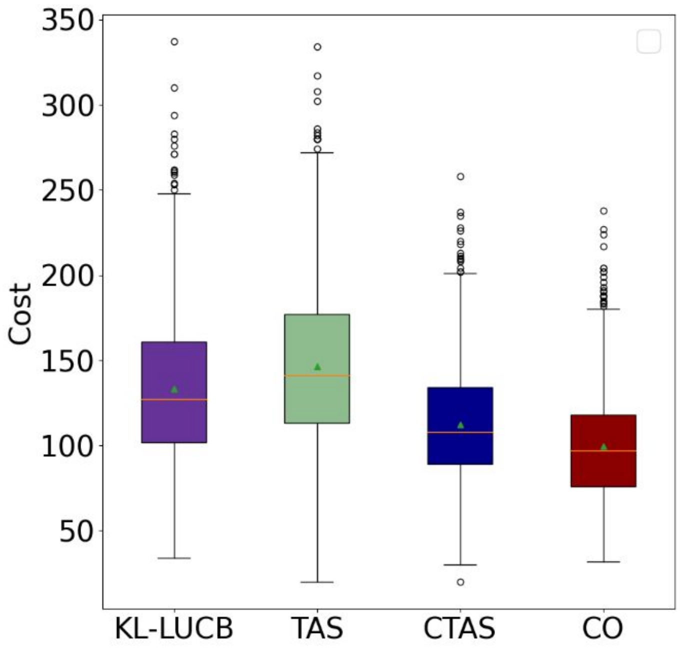
Much cheaper! ⟹

	CO	CTAS	TAS	d-LUCB
Gaussian	85	1712	2410	82
Bernoulli	58	1995	2780	60
Poisson	96	3260	4633	101

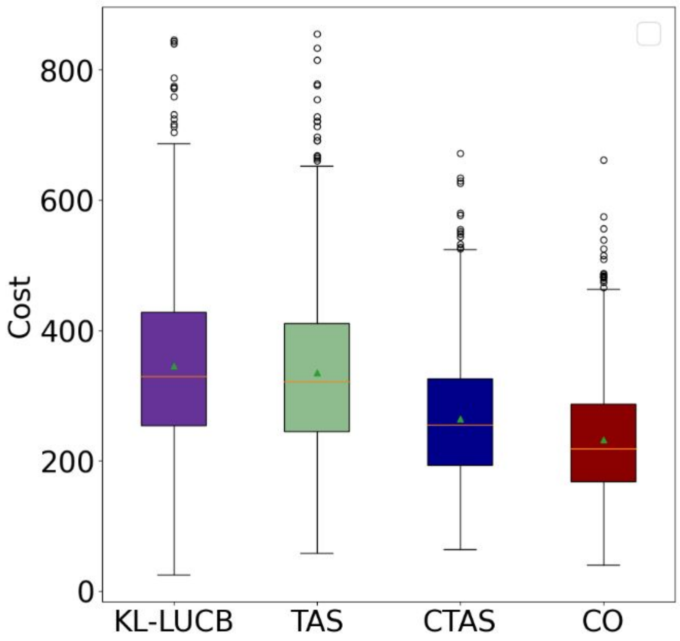
Process time (sec) over 1000 trajectories.

Kanarios, Kellen; Zhang, Qining; Ying, Lei. (2024). Cost Aware Best Arm Identification.

Results

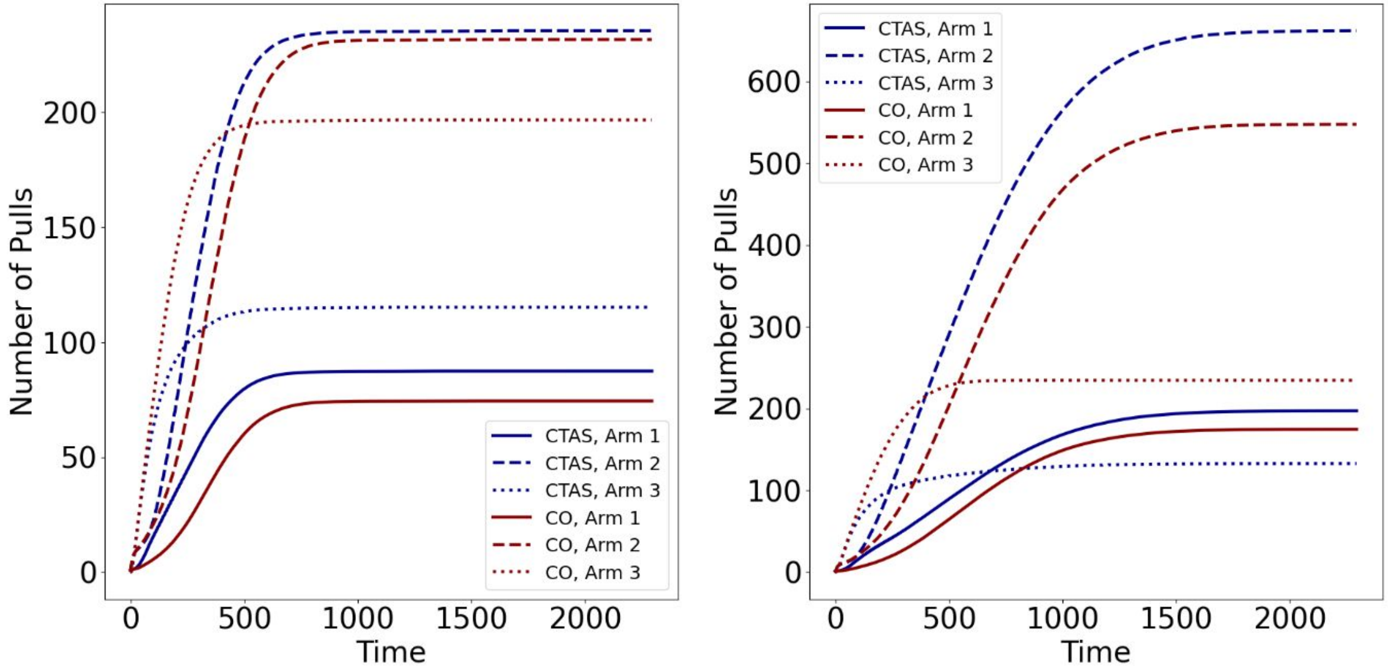


350
300
250
200
150
100
50
Cost
KL-LUCB
TAS
CTAS
CO



800
600
400
200
0
Cost
KL-LUCB
TAS
CTAS
CO

Results



Number of Pulls
Time
0
50
100
150
200
0
500
1000
1500
2000
CTAS, Arm 1
CTAS, Arm 2
CTAS, Arm 3
CO, Arm 1
CO, Arm 2
CO, Arm 3
Number of Pulls
Time
0
100
200
300
400
500
600
0
500
1000
1500
2000
CTAS, Arm 1
CTAS, Arm 2
CTAS, Arm 3
CO, Arm 1
CO, Arm 2
CO, Arm 3

Thanks!