Constrained Best Arm Identification

Wouter M. Koolen

CNI seminar, IISc, June 24, 2025

CWI and University of Twente

Warm Thanks

Tyron Lardy

Christina Katsimerou

Menu

1. Intro
2. Twist
3. Problem
4. The Lower Bound Driving Algorithm Design
5. Implementing the Interface for our Three Models
6. Achieving Asymptotic Optimality
7. Empirical Results

Intro

Motivating question

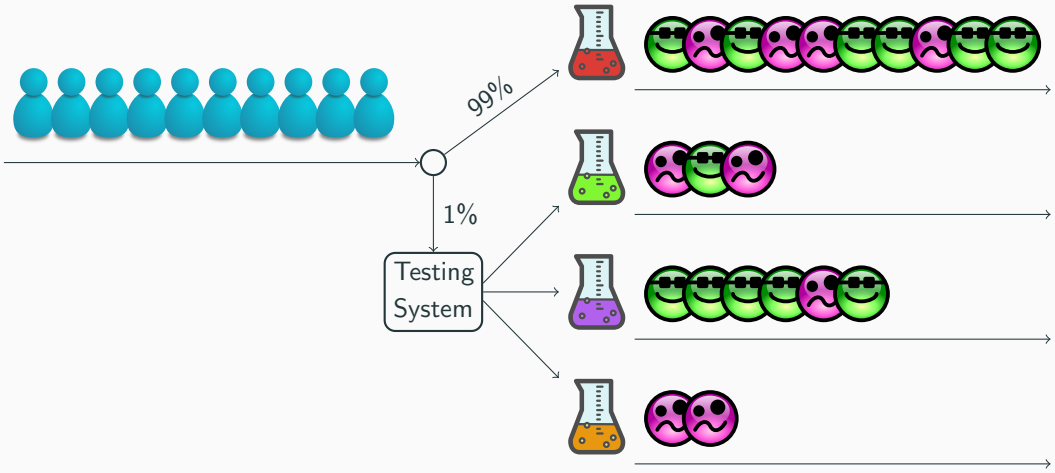
Better to switch from current system  to new version  or  or  ?

Motivating question

Better to switch from current system  to new version  or  or  ?

- A/B testing
- Adaptive clinical trial
- Best arm identification

Let's find out in production!



Best version?

Efficient asymptotically instance-optimal algorithms

Model	BAI
Gaussian reward, known covariance	(Garivier and Kaufmann, 2016)
Gaussian reward, unknown covariance	(Jourdan, Degenne, and Kaufmann, 2023)
Non-parametric reward on unit interval	(Agrawal, Juneja, and Glynn, 2020)

Twist

Toward Practical Best Arm Identification

Classical BAI is about finding the most effective arm (highest expected **reward**).
$\Rightarrow$ Arms are **univariate** distributions.

Toward Practical Best Arm Identification

Classical BAI is about finding the most effective arm (highest expected **reward**).
$\Rightarrow$ Arms are **univariate** distributions.

Often our testing task involves a constraint:

- Find most effective promotion strategy within budget constraint
- Find most effective ad bidding strategy within ROI constraint
- Find most effective treatment within safety constraint
- Find most effective code within crash percentage constraint
- ...

$\Rightarrow$ Arms are bivariate distributions.

Upgrade with Constraints

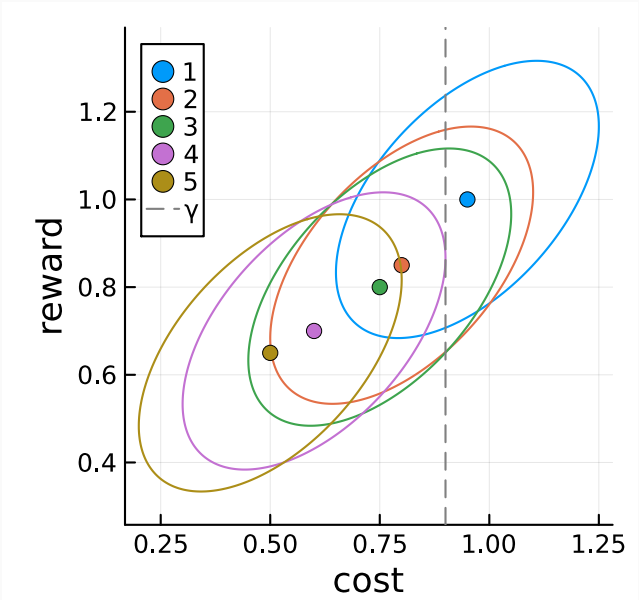
We upgrade arms to bivariate distributions on reward and cost.

CBAI: find the arm of highest expected reward among all arms with expected cost below a given threshold γ

- Reward and cost are *typically dependent*.
- Dependency structure matters! Must be learned. Must be reasoned about.
- How?

$\Rightarrow$ Crucial what we assume about the joint distribution

Which is the constrained best arm?



Problem

CBAI

An arm model $\mathcal{M}$ is a collection of distributions on $\mathbb{R}^2$. (we'll focus on three arm models)

We denote the mean of an arm $\nu \in \mathcal{M}$ by $\boldsymbol{m}(\nu) = (m_1(\nu), m_2(\nu))$.

A **bandit** with K arms from $\mathcal{M}$ is an element of $\mathcal{M}^K$.

CBAI

An arm model $\mathcal{M}$ is a collection of distributions on $\mathbb{R}^2$. (we'll focus on three arm models)

We denote the mean of an arm $\nu \in \mathcal{M}$ by $\boldsymbol{m}(\nu) = (m_1(\nu), m_2(\nu))$.

A **bandit** with K arms from $\mathcal{M}$ is an element of $\mathcal{M}^K$.

Definition

Fix threshold $\gamma \in \mathbb{R}$. The constrained best arm of bandit $\boldsymbol{\nu} \in \mathcal{M}^K$ is

$$i^*(\boldsymbol{\nu}) := \operatorname*{arg\,max}_{\substack{k \in [K] \\ m_2(\nu_k) \leq \gamma}} m_1(\nu_k)$$

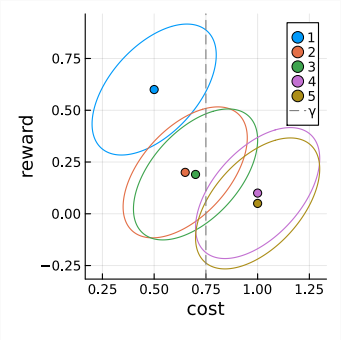
where we introduce the convention that $\arg\max \emptyset := \texttt{None}$, and assume no ties for max.

NB: i^* maps K-armed bandit to $K+1$ answers $\{1, \ldots, K, \texttt{None}\}$.

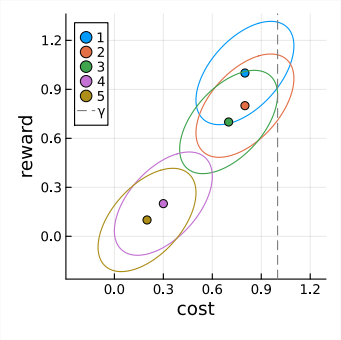
NB: only accesses bandit $\boldsymbol{\nu}$ through arm means $\boldsymbol{m}(\nu_1) \cdots \boldsymbol{m}(\nu_K)$

Practice w. CBAI definition

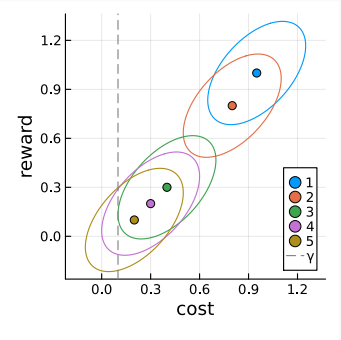
Which is the constrained best arm?



(a)



(b)



(c)

Three Models

1. Gaussian with fixed covariance $\Sigma \succeq 0$: $\mathcal{M}_{G,\Sigma} := \left\{\mathcal{N}(\boldsymbol{\mu}, \Sigma) \middle| \boldsymbol{\mu} \in \mathbb{R}^2\right\}$.
2. Gaussian with unknown covariance: $\mathcal{M}_G := \left\{\mathcal{N}(\boldsymbol{\mu}, \Sigma) \middle| \boldsymbol{\mu} \in \mathbb{R}^2, \Sigma \succeq 0\right\}$.
3. Non-parametric distributions on the unit square: $\mathcal{M}_B := \left\{P \middle| P \text{ on } [0,1]^2\right\}$.

Protocol

We work in the setting of **fixed confidence** $\delta \in (0,1)$. Fix bandit $\boldsymbol{\nu} \in \mathcal{M}^K$.

Protocol

For $t = 1, 2, \ldots, \tau$*:*

- *Learner picks an arm* $I_t \in [K]$.
- *Learner sees reward-cost pair* $(R_t, C_t) \sim \nu_{I_t}$

Learner recommends constrained best arm $\hat{\imath} \in \{1, \ldots, K, \texttt{None}\}$.

Protocol

We work in the setting of **fixed confidence** $\delta \in (0,1)$. Fix bandit $\boldsymbol{\nu} \in \mathcal{M}^K$.

Protocol

For $t = 1, 2, \ldots, \tau$*:*

- *Learner picks an arm* $I_t \in [K]$.
- *Learner sees reward-cost pair* $(R_t, C_t) \sim \nu_{I_t}$

Learner recommends constrained best arm $\hat{\imath} \in \{1, \ldots, K, \texttt{None}\}$.

Strategy for Learner specified by

- sampling rule I_t
- stopping rule τ
- recommendation rule $\hat{\imath}$

Fixed Confidence Setting

Fix $\delta \in (0,1)$. An algorithm is δ-correct if for every bandit $\boldsymbol{\nu} \in \mathcal{M}^K$

$$\mathbb{P}_{\boldsymbol{\nu}} \{\tau < \infty \text{ and } \hat{\imath} \neq i^*(\boldsymbol{\nu})\} \ \leq \ \delta.$$

Among δ-correct algorithms, we aim to **minimise the sample complexity**

$$\mathop{\mathbb{E}}_{\boldsymbol{\nu}}[\tau]$$

The Lower Bound Driving Algorithm Design

The Story from Here

- We follow the Track-and-Stop approach by Garivier and Kaufmann, 2016
  1. Prove instance-dependent sample complexity lower bound
  2. Characterise instance-optimal sampling proportions of arms
  3. Design sampling rule to match
  4. Combine with GLRT stopping and recommendation (δ-correct)
  5. $\Rightarrow$ algorithm with asymptotically optimal sample complexity
- Main ingredient that needs updating is the lower bound
- New instance-optimal sampling proportions
- And question of how to compute them

Information Theoretic Lower Bound

Theorem (Garivier and Kaufmann, 2016)

Let $\delta \in (0,1)$. For any δ-correct strategy with stopping time τ and any bandit $\boldsymbol{\nu} \in \mathcal{M}^K$,

$$\mathbb{E}_{\boldsymbol{\nu}}[\tau] \;\geq\; T^*(\boldsymbol{\nu})\, \mathrm{kl}\left(\delta \| 1-\delta\right),$$

where

$$T^*(\boldsymbol{\nu})^{-1} \;=\; \max_{\boldsymbol{w} \in \triangle_K} \min_{\substack{\boldsymbol{\nu}' \in \mathcal{M}^K \\ i^*(\boldsymbol{\nu}) \neq i^*(\boldsymbol{\nu}')}} \sum_{k=1}^{K} w_k \,\mathrm{KL}\left(\nu_k \| \nu'_k\right). \tag{1}$$

As $i^*(\boldsymbol{\nu})$ is a function of the means $\boldsymbol{m}(\nu_1) \cdots \boldsymbol{m}(\nu_K)$, we can simplify this to

$$T^*(\boldsymbol{\nu})^{-1} \;=\; \max_{\boldsymbol{w} \in \triangle_K} \min_{\substack{\boldsymbol{\lambda} \in \mathbb{R}^{K \times 2} \\ i^*(\boldsymbol{\nu}) \neq i^*(\boldsymbol{\lambda})}} \sum_{k=1}^{K} w_k \,\mathrm{KLinf}(\nu_k, \boldsymbol{\lambda}_k). \tag{2}$$

$$\text{where} \qquad \mathrm{KLinf}(\nu, \boldsymbol{\lambda}) \;\coloneqq\; \min_{\substack{\nu' \in \mathcal{M} \\ \boldsymbol{m}(\nu') = \boldsymbol{\lambda}}} \mathrm{KL}\left(\nu \| \nu'\right).$$

KLInf

$$T^*(\boldsymbol{\nu})^{-1} \;=\; \max_{\boldsymbol{w}\in\triangle_K} \min_{\substack{\boldsymbol{\lambda}\in\mathbb{R}^{K\times 2}\\ i^*(\boldsymbol{\nu})\neq i^*(\boldsymbol{\lambda})}} \sum_{k=1}^{K} w_k \,\mathrm{KLinf}(\nu_k,\boldsymbol{\lambda}_k).$$

Example

Consider a Gaussian arm $\nu = \mathcal{N}(\boldsymbol{\mu},\Sigma)$.

For Gaussians with fixed covariance Σ, i.e. $\mathcal{M}_{G,\Sigma} \;=\; \left\{\mathcal{N}(\boldsymbol{\mu},\Sigma)\middle|\boldsymbol{\mu}\in\mathbb{R}^2\right\}$,

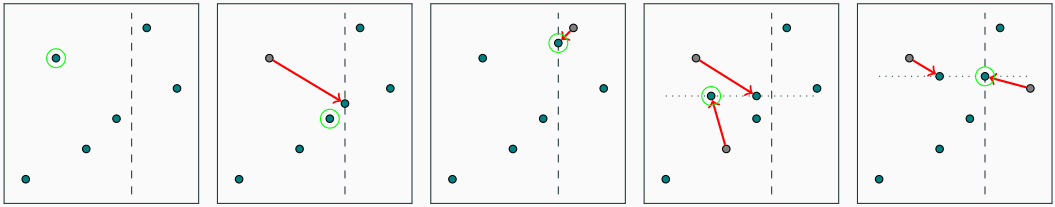
$$\mathrm{KLinf}(\nu,\boldsymbol{\lambda}) \;=\; \frac{1}{2}\left\|\boldsymbol{\mu}-\boldsymbol{\lambda}\right\|^2_{\Sigma^{-1}}$$

For Gaussians with **unknown covariance** $\mathcal{M}_G \;=\; \left\{\mathcal{N}(\boldsymbol{\mu},\Sigma)\middle|\boldsymbol{\mu}\in\mathbb{R}^2,\Sigma\succeq 0\right\}$

$$\mathrm{KLinf}(\nu,\boldsymbol{\lambda}) \;=\; \frac{1}{2}\ln\left(1+\left\|\boldsymbol{\mu}-\boldsymbol{\lambda}\right\|^2_{\Sigma^{-1}}\right)$$

Understanding the Alternative

$$T^*(\nu)^{-1} = \max_{w \in \triangle_K} \min_{\substack{\lambda \in \mathbb{R}^{K \times 2} \\ i^*(\nu) \neq i^*(\lambda)}} \sum_{k=1}^{K} w_k \, \mathrm{KLinf}(\nu_k, \lambda_k).$$



We only ever need to move two arms.

Extracting a Model-Independent Interface

1. the cost for making arm ν_j beat arm ν_i (here i can be assumed feasible)

$$c_1(\nu_i, \nu_j, w) \;:=\; \min_{\substack{\boldsymbol{\lambda}_i, \boldsymbol{\lambda}_j \in \mathbb{R}^2 \\ \lambda_{i,1} \le \lambda_{j,1} \text{ and } \lambda_{j,2} \le \gamma}} \mathsf{KLinf}(\nu_i, \boldsymbol{\lambda}_i) + w\, \mathsf{KLinf}(\nu_j, \boldsymbol{\lambda}_j). \tag{3}$$

2. the cost for **changing the feasibility** status of an arm ν

$$c_2(\nu) \;:=\; \begin{cases} \min_{\substack{\boldsymbol{\lambda} \in \mathbb{R}^2 \\ \lambda_2 \ge \gamma}} \mathsf{KLinf}(\nu, \boldsymbol{\lambda}) & \text{if } m_2(\nu) \le \gamma \\ \min_{\substack{\boldsymbol{\lambda} \in \mathbb{R}^2 \\ \lambda_2 \le \gamma}} \mathsf{KLinf}(\nu, \boldsymbol{\lambda}) & \text{if } m_2(\nu) > \gamma \end{cases}.$$

In terms of this interface, our problem simplifies to

$$T^*(\boldsymbol{\nu})^{-1} \;=\; \max_{\boldsymbol{w} \in \triangle_K} \begin{cases} \min\left\{ \min_{j \neq i^*} \; w_{i^*} c_1\left(\nu_{i^*}, \nu_j, \frac{w_j}{w_{i^*}}\right), w_{i^*} c_2(\nu_{i^*}) \right\} & \text{if } i^* \neq \texttt{None}, \\ \min_{j \in [K]} \; w_j c_2(\nu_j) & \text{if } i^* = \texttt{None}. \end{cases} \tag{4}$$

Characterisation of Sample Complexity

Theorem

Fix bandit $\boldsymbol{\nu} \in \mathcal{M}^K$. Let $i^* := i^*(\boldsymbol{m})$. For all $i \in [K]$, we have

$$T^*(\boldsymbol{\nu}) \;=\; \begin{cases} \frac{\sum_{j=1}^K \tilde{w}_j(\tilde{C}^*)}{\tilde{C}^*} \\ \sum_{j=1}^K c_2(\nu_j)^{-1} \end{cases} \qquad \text{and} \qquad w_i^*(\boldsymbol{\nu}) \;=\; \begin{cases} \frac{\tilde{w}_i(\tilde{C}^*)}{\sum_{j=1}^K \tilde{w}_j(\tilde{C}^*)} & \text{if } i^* \neq \texttt{None} \\ \frac{c_2(\nu_i)^{-1}}{\sum_{j=1}^K c_2(\nu_j)^{-1}} & \text{if } i^* = \texttt{None} \end{cases}$$

where $\tilde{w}_{i^*}(\tilde{C}) := 1$, and for each sub-optimal $j \neq i^*$, $\tilde{w}_j(\tilde{C})$ is the unique solution to w in

$$c_1(\nu_{i^*}, \nu_j, w) \;=\; \tilde{C}, \tag{5}$$

and $\tilde{C}^*$ is the unique solution for $\tilde{C}$ in

$$\sum_{j \neq i^*} \frac{c_{1,i^*}\big(\nu_{i^*}, \nu_j, \tilde{w}_j(\tilde{C})\big)}{c_{1,j}\big(\nu_{i^*}, \nu_j, \tilde{w}_j(\tilde{C})\big)} \;=\; 1 \tag{6}$$

clamped to the interval $\big[0, c_2(\nu_{i^*})\big]$.

Efficient Computation

One outer binary search to compute $\tilde{C}$.

One inner binary search per arm to compute $\tilde{w}_j(C)$.

Same computational cost as (Garivier and Kaufmann, 2016) for the oracle weights in BAI.

It remains to implement c_1 and c_2.

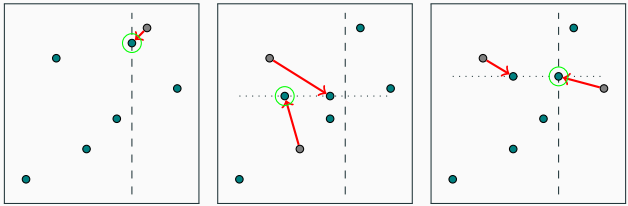
Implementing the Interface for our Three Models

Recall our three models

1. Gaussian with fixed covariance $\Sigma \succeq 0$: $\mathcal{M}_{G,\Sigma} \coloneqq \left\{\mathcal{N}(\boldsymbol{\mu}, \Sigma) \middle| \boldsymbol{\mu} \in \mathbb{R}^2\right\}$.
2. Gaussian with unknown covariance: $\mathcal{M}_G \coloneqq \left\{\mathcal{N}(\boldsymbol{\mu}, \Sigma) \middle| \boldsymbol{\mu} \in \mathbb{R}^2, \Sigma \succeq 0\right\}$.
3. Non-parametric distributions on the unit square: $\mathcal{M}_B \coloneqq \left\{P \middle| P \text{ on } [0,1]^2\right\}$.

Here we implement the most interesting function form the interface

$$c_1(\nu_i, \nu_j, w) \coloneqq \min_{\substack{\boldsymbol{\lambda}_i, \boldsymbol{\lambda}_j \in \mathbb{R}^2 \\ \lambda_{i,1} \leq \lambda_{j,1} \text{ and } \lambda_{j,2} \leq \gamma}} \mathrm{KLinf}(\nu_i, \boldsymbol{\lambda}_i) + w\,\mathrm{KLinf}(\nu_j, \boldsymbol{\lambda}_j).$$



Gaussian with Known Covariance Σ

Theorem

Fix bivariate $\nu_i = \mathcal{N}(\boldsymbol{\mu}_i, \Sigma)$ *and* $\nu_j = \mathcal{N}(\boldsymbol{\mu}_j, \Sigma)$ *with* $i^*(\{\boldsymbol{\mu}_i, \boldsymbol{\mu}_j\}) = i$*, then*

$$c_1(\nu_i, \nu_j, w) = \begin{cases} \frac{w(\mu_{j,2}-\gamma)^2}{2\Sigma_{22}} & \text{if } \mu_{j,1} - \frac{\Sigma_{12}}{\Sigma_{22}}(\mu_{j,2}-\gamma)_+ \geq \mu_{i,1} \\ \frac{(\mu_{j,1}-\mu_{i^*,1})^2}{2\Sigma_{11}(1+\frac{1}{w})} & \text{if } \mu_{j,2} + \frac{\frac{1}{w}\Sigma_{12}}{\Sigma_{i,11}+\frac{1}{w}\Sigma_{11}}(\mu_{i,1}-\mu_{j,1}) \leq \gamma \\ \frac{w\Sigma_{11}(\gamma-\mu_{j,2})^2+|\Sigma|\left\|\begin{matrix}\mu_{j,1}-\mu_{i^*,1}\\ \mu_{j,2}-\gamma\end{matrix}\right\|^2_{\Sigma^{-1}}}{2\left(\Sigma_{11}\Sigma_{22}+|\Sigma|\frac{1}{w}\right)} & \text{else.} \end{cases}$$

Closed form, $O(1)$ per arm.

Gaussian with Unknown Covariance

Theorem

Fix bivariate $\nu_i = \mathcal{N}(\boldsymbol{\mu}_i, \Sigma_i)$ *and* $\nu_j = \mathcal{N}(\boldsymbol{\mu}_j, \Sigma_j)$. *Abbreviating* $\ell(x) := \frac{1}{2}\ln(1+x)$,

$$c_1(\nu_i, \nu_j, w) \;=\; \min_{\theta \in \mathbb{R}} \; \ell\left(\frac{(\mu_{i,1}-\theta)_+^2}{\Sigma_{i,11}}\right) + w \begin{cases} 0 & \text{if } \mu_{i,2} \;\le\; \gamma \text{ and } \mu_{j,1} \;\ge\; \theta \\ \ell\left(\frac{(\mu_{j,2}-\gamma)_+^2}{\Sigma_{j,22}}\right) & \text{if } \mu_{j,1} - \frac{\Sigma_{j,12}}{\Sigma_{j,22}}(\mu_{j,2}-\gamma)_+ \;\ge\; \theta \\ \ell\left(\frac{(\mu_{j,1}-\theta)_-^2}{\Sigma_{j,11}}\right) & \text{if } \mu_{j,2} + \frac{\Sigma_{j,12}}{\Sigma_{j,11}}(\mu_{j,1}-\theta)_- \;\le\; \gamma \\ \mathrm{KLinf}(\nu_j, (\theta, \gamma)) & \text{else.} \end{cases}$$

This is the minimum (in θ) of four sum-of-log-of-one-plus-square. Cancelling the derivative results in a cubic equation. Even with careful tracking of **case jurisdictions**, $O(1)$ per arm.

Non-parametric distributions on the unit square

Theorem

Let ν_i, ν_j be bivariate distributions on $[0,1]^2$. Then

$$c_1(\nu_i, \nu_j, w) \;=\; \max_{\substack{\boldsymbol{b}\in(\star)\\ b_3\geq 0\geq b_2}} \; \mathop{\mathbb{E}}_{\nu_i}[\ln(1 - w(b_1 + b_2 R))] + w \mathop{\mathbb{E}}_{\nu_j}[\ln(1 + b_1 + b_2 R + b_3(C - \gamma))]$$

where $(\star)$ ensures that the argument of the log is positive for all (x_1, x_2) in the unit square.

The constraints on $\boldsymbol{b}$ are a polyhedron in 3 variables with six faces.

For ν_i, ν_j supported on n points, this takes time $O(n)$ with e.g. Ellipsoid.

Achieving Asymptotic Optimality

Steps to a full Algorithm

We saw the calculation of the characteristic time $T^*(\nu)$ and the **oracle weights** $w^*(\nu)$.

The rest follows the track-and-stop (TaS) framework.

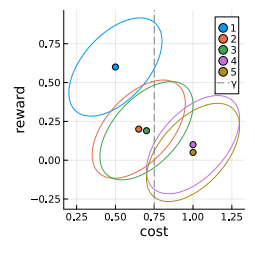
- Empirical plug-in estimate of the bandit
- GLR stopping rule
- Empirical answer recommendation

Theorem

TaS is asymptotically optimal, i.e. $\lim_{\delta \to 0} \frac{\mathbb{E}_\nu[\tau_\delta]}{\ln \frac{1}{\delta}} = T^*(\nu)$.

Empirical Results

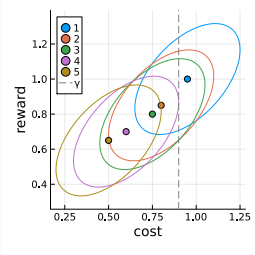
Sample Complexity



Easy

10.5, 48.3

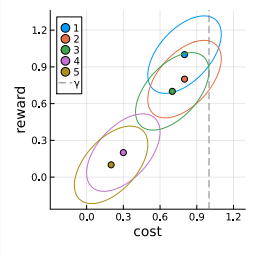
36, 24, 21, 10, 09



Hard

410.4, 1890.0

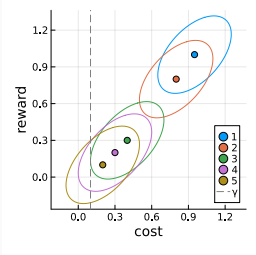
18, 39, 39, 03, 01



All feasible

26.5, 122.0

40, 39, 14, 04, 03



None feasible

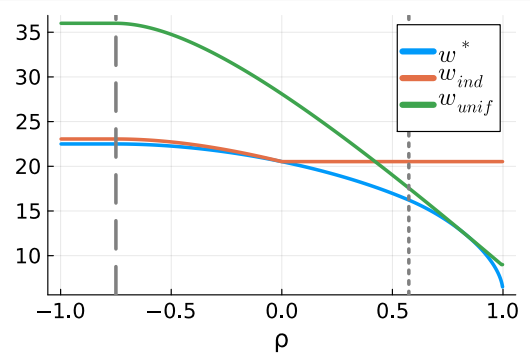
29.3, 134.9

03, 04, 10, 19, 65

Instance	TaS-EV	TaS	Oracle	Uniform	TopTwo-TCI	Racing	TaS-1d
Easy	83.4± 0.3	67.5± 0.3	89.6± 0.4	136.4± 0.6	68.8± 0.6	100.8± 7.8	96.6± 0.5
Hard	2772.3± 56.4	2619.9± 54.2	4225.4± 59.4	5498.9±129.8	2859.9± 54.9	3465.4± 59.3	4815.9±101.6
All feasible	210.9± 2.8	180.9± 2.6	229.7± 2.3	354.0± 4.8	174.6± 2.4	230.0± 2.7	186.4± 2.3
None feasible	273.0± 4.6	200.2± 3.7	270.1± 3.1	576.0± 13.4	241.9± 5.4	219.1± 3.4	3293.4± 84.4

The Impact of Dependency

We study the following two-arm problem $\boldsymbol{\nu}_\rho$ in the fixed covariance Gaussian model as a **function of correlation** $\rho \in [-1, 1]$: $\gamma = \frac{2}{3}$, $\mu_1 = (0, 0)$, $\mu_2 = \left(-\frac{1}{4}, 1\right)$, cost and reward each have variance 1, and the correlation between them is ρ.



Sample complexity $T^*(\boldsymbol{\nu}_\rho)$.

Conclusion

Results: Efficient asymptotically instance-optimal algorithms

Model	BAI (1d)	CBAI (2d)
Gaussian, known covariance	(Garivier and Kaufmann, 2016)	Here
Gaussian, unknown covariance	(Jourdan, Degenne, and Kaufmann, 2023)	Here
Non-parametric on hypercube	(Agrawal, Juneja, and Glynn, 2020)	Here

Conclusion

We motivated the constrained best arm identification problem.

This necessitated going **bivariate** (reward and cost).

We developed asymptotically optimal algorithms for different model assumptions.

We extracted a generic **interface** for analysis and computation

And implemented it efficiently for the three models

The method works in practice.

Let's talk!