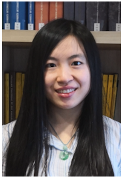
Optimal Clustering with Bandit Feedback

https://arxiv.org/abs/2202.04294

Junwen Yang, Zixin Zhong, Vincent Tan



National University of Singapore

Stanford RL Seminar

May, 24 2022

Outline

Outline

Motivation

Clustering

- The task of partitioning a set of items into smaller clusters
- One of the most fundamental tasks in machine learning
- Numerous algorithms proposed (e.g., K-means and spectral clustering)

Motivation

Clustering

- The task of partitioning a set of items into smaller clusters
- One of the most fundamental tasks in machine learning
- Numerous algorithms proposed (e.g., K-means and spectral clustering)

Challenges

- Measurement noise
- Sequential and adaptive data collection

Online Clustering with Bandit Feedback (Informal)

High-Level Description of our Problem

- An online variant of the classical offline clustering problem

Online Clustering with Bandit Feedback (Informal)

High-Level Description of our Problem

- An online variant of the classical offline clustering problem
- Bandit feedback: at each time step, the agent only observes a noisy measurement on the selected arm (or item)

Online Clustering with Bandit Feedback (Informal)

High-Level Description of our Problem

- An online variant of the classical offline clustering problem
- Bandit feedback: at each time step, the agent only observes a noisy measurement on the selected arm (or item)
- Pull arms adaptively, so as to minimize the expected number of total arm pulls it takes to correctly partition the given arm set with a given high probability

Applications in Digital Marketing

Sequential Collection of Customer Feedback

Customer feedback on certain products are collected in an online manner and always accompanied by random or systematic noise.

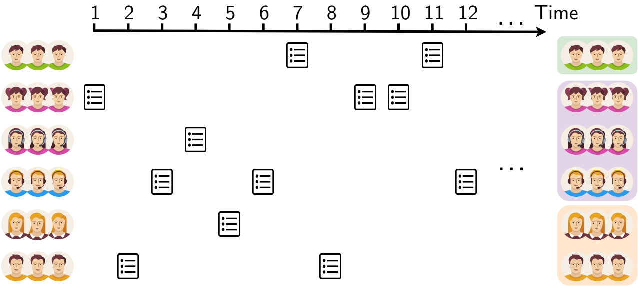


Figure: Example involving partitioning 6 sub-groups of customers into 3 market segments with bandit feedback.

Outline

A Bandit Feedback Model with Cluster Structure

Mathematical Model I

- The arm set $\mathcal{A} = [M]$ can be partitioned into K disjoint nonempty clusters ($K < M$).

A Bandit Feedback Model with Cluster Structure

Mathematical Model I

- The arm set $\mathcal{A} = [M]$ can be partitioned into K disjoint nonempty clusters ($K < M$).
- In the same cluster, the arms share the same d-dimensional mean vector (center).

A Bandit Feedback Model with Cluster Structure

Mathematical Model I

- The arm set $\mathcal{A} = [M]$ can be partitioned into K disjoint nonempty clusters ($K < M$).
- In the same cluster, the arms share the same d-dimensional mean vector (center).
- An instance of cluster bandits can be fully characterized by a pair $(c, \mathcal{U})$, where $c = [c_1, c_2, \ldots, c_M] \in [K]^M$ consists of the cluster indices of the arms and $\mathcal{U} = [\mu(1), \mu(2), \ldots, \mu(K)] \in \mathbb{R}^{d \times K}$ represents the K centers of the clusters.

A Bandit Feedback Model with Cluster Structure

Mathematical Model I

- The arm set $\mathcal{A} = [M]$ can be partitioned into K disjoint nonempty clusters ($K < M$).
- In the same cluster, the arms share the same d-dimensional mean vector (center).
- An instance of cluster bandits can be fully characterized by a pair $(c, \mathcal{U})$, where $c = [c_1, c_2, \ldots, c_M] \in [K]^M$ consists of the cluster indices of the arms and $\mathcal{U} = [\mu(1), \mu(2), \ldots, \mu(K)] \in \mathbb{R}^{d \times K}$ represents the K centers of the clusters.
- Only consider partitioning the instances in which the mean vectors for different clusters are distinct.

A Bandit Feedback Model with Cluster Structure

Mathematical Model II

- The cluster indices of the M arms: $c = [c_1, c_2, \ldots, c_M] \in [K]^M$

A Bandit Feedback Model with Cluster Structure

Mathematical Model II

- The cluster indices of the M arms: $c = [c_1, c_2, \ldots, c_M] \in [K]^M$
- The K centers of the clusters: $\mathcal{U} = [\mu(1), \mu(2), \ldots, \mu(K)] \in \mathbb{R}^{d \times K}$

A Bandit Feedback Model with Cluster Structure

Mathematical Model II

- The cluster indices of the M arms: $c = [c_1, c_2, \ldots, c_M] \in [K]^M$
- The K centers of the clusters: $\mathcal{U} = [\mu(1), \mu(2), \ldots, \mu(K)] \in \mathbb{R}^{d \times K}$
- At each time t, the agent selects an arm A_t from the arm set $\mathcal{A}$, and then observes an noisy measurement on the mean vector of A_t, i.e.,

$$X_t = \mu(c_{A_t}) + \eta_t$$

  where $\eta_t \in \mathbb{R}^d$ is independent noise, following the standard d-dimensional Gaussian distribution $\mathcal{N}(0, \mathrm{I}_d)$.

A Bandit Feedback Model with Cluster Structure

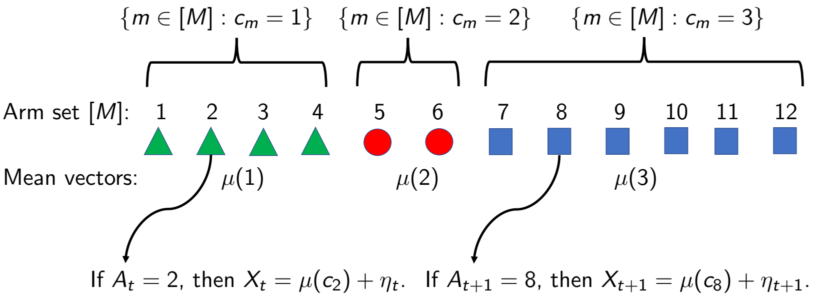


Figure: Online clustering with bandit feedback with $K = 3$ and $M = 12$.

The Equivalences of Partitions and Instances

- Representation of a partition c or an instance $(c, \mathcal{U})$ is not unique.

The Equivalences of Partitions and Instances

- Representation of a partition c or an instance $(c, \mathcal{U})$ is not unique.
- For a permutation σ on $[K]$, let

$$\sigma(c) := [\sigma(c_1), \sigma(c_2), \ldots, \sigma(c_M)]$$

and

$$\sigma(\mathcal{U}) := [\mu(\sigma(1)), \mu(\sigma(2)), \ldots, \mu(\sigma(K))].$$

Similarly, $\sigma(c, \mathcal{U}) := (\sigma(c), \sigma(\mathcal{U}))$.

The Equivalences of Partitions and Instances

- Representation of a partition c or an instance $(c, \mathcal{U})$ is not unique.
- For a permutation σ on $[K]$, let

$$\sigma(c) := [\sigma(c_1), \sigma(c_2), \ldots, \sigma(c_M)]$$

and

$$\sigma(\mathcal{U}) := [\mu(\sigma(1)), \mu(\sigma(2)), \ldots, \mu(\sigma(K))].$$

Similarly, $\sigma(c, \mathcal{U}) := (\sigma(c), \sigma(\mathcal{U}))$.

- For two partitions c and c', if there exists a permutation σ on $[K]$ such that $c = \sigma(c')$, then we write $c \sim c'$.

The Equivalences of Partitions and Instances

- Representation of a partition c or an instance $(c, \mathcal{U})$ is not unique.
- For a permutation σ on $[K]$, let

  $$\sigma(c) := [\sigma(c_1), \sigma(c_2), \ldots, \sigma(c_M)]$$

  and

  $$\sigma(\mathcal{U}) := [\mu(\sigma(1)), \mu(\sigma(2)), \ldots, \mu(\sigma(K))].$$

  Similarly, $\sigma(c, \mathcal{U}) := (\sigma(c), \sigma(\mathcal{U}))$.
- For two partitions c and c', if there exists a permutation σ on $[K]$ such that $c = \sigma(c')$, then we write $c \sim c'$.
- For two instances $(c, \mathcal{U})$ and $(c', \mathcal{U}')$, if $\mu(c_m) = \mu'(c'_m)$ for all $m \in [M]$, then we write $(c, \mathcal{U}) \sim (c', \mathcal{U}')$.

Online Clustering with Bandit Feedback

What is an Online Algorithm?

In the fixed-confidence setting where a confidence level $\delta \in (0, 1)$ is given, the agent uses an *online algorithm* π:

Online Clustering with Bandit Feedback

What is an Online Algorithm?

In the fixed-confidence setting where a confidence level $\delta \in (0, 1)$ is given, the agent uses an *online algorithm* π:

- to decide the arm A_t to pull at each time step t, (*sampling rule*)

Online Clustering with Bandit Feedback

What is an Online Algorithm?

In the fixed-confidence setting where a confidence level $\delta \in (0,1)$ is given, the agent uses an *online algorithm* π:

- to decide the arm A_t to pull at each time step t, (*sampling rule*)
- to choose a time τ_δ to stop pulling arms, (*stopping rule*)

Online Clustering with Bandit Feedback

What is an Online Algorithm?

In the fixed-confidence setting where a confidence level $\delta \in (0, 1)$ is given, the agent uses an *online algorithm* π:

- to decide the arm A_t to pull at each time step t, (*sampling rule*)
- to choose a time τ_δ to stop pulling arms, (*stopping rule*)
- and to recommend a *correct* partition c^{out} of the arm set $[M]$ (i.e., $c^{\mathrm{out}} \sim c$) with a probability of at least $1 - \delta$ (*recommendation rule*)

in the smallest expected number of time steps.

Online Clustering with Bandit Feedback

What is an Online Algorithm?

In the fixed-confidence setting where a confidence level $\delta \in (0,1)$ is given, the agent uses an *online algorithm* π:

- to decide the arm A_t to pull at each time step t, (*sampling rule*)
- to choose a time τ_δ to stop pulling arms, (*stopping rule*)
- and to recommend a *correct* partition c^{out} of the arm set $[M]$ (i.e., $c^{\mathrm{out}} \sim c$) with a probability of at least $1-\delta$ (*recommendation rule*)

in the smallest expected number of time steps.

Performance Metric

$$\min_{\pi} \mathbb{E}[\tau_\delta] \quad \text{s.t.} \quad \underbrace{\Pr(\tau_\delta < \infty) = 1 \text{ and } \Pr(c^{\mathrm{out}} \not\sim c) \le \delta}_{\delta\text{-PAC}}$$

Other Notations

- $\mathcal{P}_N := \{x \in [0,1]^N : \|x\|_1 = 1\}$ denotes the probability simplex in $\mathbb{R}^N$ while $\mathcal{P}_N^+ := \{x \in (0,1)^N : \|x\|_1 = 1\}$ denotes the open simplex.

Other Notations

- $\mathcal{P}_N := \{x \in [0,1]^N : \|x\|_1 = 1\}$ denotes the probability simplex in $\mathbb{R}^N$ while $\mathcal{P}_N^+ := \{x \in (0,1)^N : \|x\|_1 = 1\}$ denotes the open simplex.
- For two partitions c and c', let

$$d_{\mathrm{H}}(c, c') := \sum_{m=1}^{M} \mathbb{1}\{c_m \neq c'_m\}.$$

Other Notations

- $\mathcal{P}_N := \{x \in [0,1]^N : \|x\|_1 = 1\}$ denotes the probability simplex in $\mathbb{R}^N$ while $\mathcal{P}_N^+ := \{x \in (0,1)^N : \|x\|_1 = 1\}$ denotes the open simplex.
- For two partitions c and c', let

$$d_{\mathrm{H}}(c, c') := \sum_{m=1}^{M} \mathbb{1}\{c_m \neq c'_m\}.$$

- For any $a, b \in (0,1)$, the binary relative entropy between Bernoulli distributions with means a and b, is denoted as

$$d_{\mathrm{KL}}(a, b) := a \log \frac{a}{b} + (1-a) \log \frac{1-a}{1-b}.$$

Other Notations

- $\mathcal{P}_N := \{x \in [0,1]^N : \|x\|_1 = 1\}$ denotes the probability simplex in $\mathbb{R}^N$ while $\mathcal{P}_N^+ := \{x \in (0,1)^N : \|x\|_1 = 1\}$ denotes the open simplex.
- For two partitions c and c', let

$$d_{\mathrm{H}}(c, c') := \sum_{m=1}^{M} \mathbb{1}\{c_m \neq c'_m\}.$$

- For any $a, b \in (0,1)$, the binary relative entropy between Bernoulli distributions with means a and b, is denoted as

$$d_{\mathrm{KL}}(a, b) := a \log \frac{a}{b} + (1-a) \log \frac{1-a}{1-b}.$$

- $i^* = \arg\min_{i \in A} f(i)$ refers to the minimum index in the set $\{i \in A : f(i) = \min_{j \in A} f(j)\}$.

Outline

The Change-of-measure Argument

- The key idea dates back to Chernoff (1959).

The Change-of-measure Argument

- The key idea dates back to Chernoff (1959).
- Ubiquitous in showing various lower bounds in bandit problems (e.g., regret minimization (Lai and Robbins, 1985; Lattimore and Szepesvári, 2017), pure exploration (Kaufmann et al., 2016; Vaidhiyan and Sundaresan, 2017)).

The Change-of-measure Argument

- The key idea dates back to Chernoff (1959).
- Ubiquitous in showing various lower bounds in bandit problems (e.g., regret minimization (Lai and Robbins, 1985; Lattimore and Szepesvári, 2017), pure exploration (Kaufmann et al., 2016; Vaidhiyan and Sundaresan, 2017)).
- Probabilities of an event under different probability measures are related via the KL-divergence between the two measures: $\forall\, A \subset \Omega$,

$$\mathbb{P}(A) + \mathbb{Q}(A^c) \geq \frac{1}{2} \exp\big(- D(\mathbb{P} \| \mathbb{Q})\big) \qquad D(\mathbb{P}\|\mathbb{Q}) := \int_\Omega \mathrm{d}\mathbb{P} \log \frac{\mathrm{d}\mathbb{P}}{\mathrm{d}\mathbb{Q}}.$$

The Change-of-measure Argument

- The key idea dates back to Chernoff (1959).
- Ubiquitous in showing various lower bounds in bandit problems (e.g., regret minimization (Lai and Robbins, 1985; Lattimore and Szepesvári, 2017), pure exploration (Kaufmann et al., 2016; Vaidhiyan and Sundaresan, 2017)).
- Probabilities of an event under different probability measures are related via the KL-divergence between the two measures: $\forall A \subset \Omega$,

$$\mathbb{P}(A) + \mathbb{Q}(A^c) \geq \frac{1}{2} \exp\left(-D(\mathbb{P}\|\mathbb{Q})\right) \qquad D(\mathbb{P}\|\mathbb{Q}) := \int_\Omega \mathrm{d}\mathbb{P} \log \frac{\mathrm{d}\mathbb{P}}{\mathrm{d}\mathbb{Q}}.$$

- For any fixed instance $(c, \mathcal{U})$, we define

$$\mathrm{Alt}(c) := \{(c', \mathcal{U}') : c'' \neq c \text{ for any } (c'', \mathcal{U}'') \sim (c', \mathcal{U}')\},$$

the set of alternative instances where c is not a correct partition.

Lower Bound

Theorem 1

For a fixed confidence level $\delta \in (0,1)$ and instance $(c,\mathcal{U})$, any δ-PAC online clustering algorithm satisfies

$$\mathbb{E}[\tau_\delta] \geq d_{\mathrm{KL}}(\delta, 1-\delta) D^*(c,\mathcal{U})$$

where

$$D^*(c,\mathcal{U}) := \left(\frac{1}{2} \sup_{\lambda \in \mathcal{P}_M} \inf_{(c',\mathcal{U}') \in \mathrm{Alt}(c)} \sum_{m=1}^{M} \lambda_m \|\mu(c_m) - \mu'(c'_m)\|^2\right)^{-1}.$$

Furthermore,

$$\liminf_{\delta \to 0} \frac{\mathbb{E}[\tau_\delta]}{\log(1/\delta)} \geq D^*(c,\mathcal{U}).$$

Hardness Parameter

The Hardness Parameter

$$D^*(c,\mathcal{U}) := \left(\frac{1}{2} \sup_{\lambda \in \mathcal{P}_M} \inf_{(c',\mathcal{U}') \in \mathrm{Alt}(c)} \sum_{m=1}^{M} \lambda_m \|\mu(c_m) - \mu'(c'_m)\|^2\right)^{-1}$$

Hardness Parameter

The Hardness Parameter

$$D^*(c,\mathcal{U}) := \left(\frac{1}{2} \sup_{\lambda \in \mathcal{P}_M} \inf_{(c',\mathcal{U}') \in \mathrm{Alt}(c)} \sum_{m=1}^{M} \lambda_m \|\mu(c_m) - \mu'(c'_m)\|^2\right)^{-1}$$

Remarks

- Any $\lambda \in \mathcal{P}_M$ can be understood as the proportion of arm pulls.

Hardness Parameter

The Hardness Parameter

$$D^*(c,\mathcal{U}) := \left(\frac{1}{2}\sup_{\lambda\in\mathcal{P}_M}\inf_{(c',\mathcal{U}')\in\mathrm{Alt}(c)}\sum_{m=1}^{M}\lambda_m\|\mu(c_m)-\mu'(c'_m)\|^2\right)^{-1}$$

Remarks

- Any $\lambda \in \mathcal{P}_M$ can be understood as the proportion of arm pulls.
- Wish to find the optimal proportion of arm pulls to distinguish the instance c from the most confusing alternative instances in $\mathrm{Alt}(c)$.

Hardness Parameter

The Hardness Parameter

$$D^*(c,\mathcal{U}) := \left(\frac{1}{2} \sup_{\lambda \in \mathcal{P}_M} \inf_{(c',\mathcal{U}') \in \mathrm{Alt}(c)} \sum_{m=1}^{M} \lambda_m \|\mu(c_m) - \mu'(c'_m)\|^2 \right)^{-1}$$

Remarks

- Any $\lambda \in \mathcal{P}_M$ can be understood as the proportion of arm pulls.
- Wish to find the optimal proportion of arm pulls to distinguish the instance c from the most confusing alternative instances in $\mathrm{Alt}(c)$.
- With the knowledge of the instance $(c,\mathcal{U})$, the optimization problem naturally reveals the optimal sampling rule, which will be the basic idea behind the design of our sampling rule.

Two Optimization Problems

Key Optimization Problems

Problem (SupInf): $$\sup_{\lambda \in \mathcal{P}_M} \inf_{(c', \mathcal{U}') \in \mathrm{Alt}(c)} \sum_{m=1}^{M} \lambda_m \|\mu(c_m) - \mu'(c'_m)\|^2$$

Problem (InnerInf): $$\inf_{(c', \mathcal{U}') \in \mathrm{Alt}(c)} \sum_{m=1}^{M} \lambda_m \|\mu(c_m) - \mu'(c'_m)\|^2$$

Two Optimization Problems

Key Optimization Problems

Problem (SupInf): $$\sup_{\lambda \in \mathcal{P}_M} \inf_{(c',\mathcal{U}') \in \mathrm{Alt}(c)} \sum_{m=1}^{M} \lambda_m \|\mu(c_m) - \mu'(c'_m)\|^2$$

Problem (InnerInf): $$\inf_{(c',\mathcal{U}') \in \mathrm{Alt}(c)} \sum_{m=1}^{M} \lambda_m \|\mu(c_m) - \mu'(c'_m)\|^2$$

Can these be solved "tractably"?

- These optimization problems in their original form appear to be intractable.

Two Optimization Problems

Key Optimization Problems

Problem (SupInf): $$\sup_{\lambda \in \mathcal{P}_M} \inf_{(c',\mathcal{U}') \in \mathrm{Alt}(c)} \sum_{m=1}^{M} \lambda_m \|\mu(c_m) - \mu'(c'_m)\|^2$$

Problem (InnerInf): $$\inf_{(c',\mathcal{U}') \in \mathrm{Alt}(c)} \sum_{m=1}^{M} \lambda_m \|\mu(c_m) - \mu'(c'_m)\|^2$$

Can these be solved "tractably"?

- These optimization problems in their original form appear to be intractable.
- The definition of $\mathrm{Alt}(c)$ is combinatorial and the number of instances in it is obviously infinite.

Two Optimization Problems

Key Optimization Problems

Problem (SupInf): $$\sup_{\lambda \in \mathcal{P}_M} \inf_{(c',\mathcal{U}') \in \mathrm{Alt}(c)} \sum_{m=1}^{M} \lambda_m \|\mu(c_m) - \mu'(c'_m)\|^2$$

Problem (InnerInf): $$\inf_{(c',\mathcal{U}') \in \mathrm{Alt}(c)} \sum_{m=1}^{M} \lambda_m \|\mu(c_m) - \mu'(c'_m)\|^2$$

Can these be solved "tractably"?

- These optimization problems in their original form appear to be intractable.
- The definition of $\mathrm{Alt}(c)$ is combinatorial and the number of instances in it is obviously infinite.
- For a fixed number of clusters K, the total number of possible partitions grows asymptotically as $K^M/K!$

A Combinatorial Property of Problem (InnerInf)

Lemma 2

For any $\lambda \in \mathcal{P}_M$ *and* $(c, \mathcal{U})$,

$$\inf_{(c',\mathcal{U}')\in \mathrm{Alt}(c)} \sum_{m=1}^{M} \lambda_m \|\mu(c_m) - \mu'(c'_m)\|^2 = \inf_{\substack{(c',\mathcal{U}')\in \mathrm{Alt}(c):\\ d_{\mathrm{H}}(c',c)=1}} \sum_{m=1}^{M} \lambda_m \|\mu(c_m) - \mu'(c'_m)\|^2.$$

A Combinatorial Property of Problem (InnerInf)

Lemma 2

For any $\lambda \in \mathcal{P}_M$ and $(c, \mathcal{U})$,

$$\inf_{(c',\mathcal{U}')\in \mathrm{Alt}(c)} \sum_{m=1}^{M} \lambda_m \|\mu(c_m) - \mu'(c'_m)\|^2 = \inf_{\substack{(c',\mathcal{U}')\in \mathrm{Alt}(c): \\ d_{\mathrm{H}}(c',c)=1}} \sum_{m=1}^{M} \lambda_m \|\mu(c_m) - \mu'(c'_m)\|^2.$$

- Instead of considering all the alternative instances in $\mathrm{Alt}(c)$, Lemma 2 shows it suffices to consider the instances whose partitions have a Hamming distance of 1 from the given partition c.

A Combinatorial Property of Problem (InnerInf)

Lemma 2

For any $\lambda \in \mathcal{P}_M$ *and* $(c, \mathcal{U})$,

$$\inf_{(c',\mathcal{U}')\in \mathrm{Alt}(c)} \sum_{m=1}^{M} \lambda_m \|\mu(c_m) - \mu'(c'_m)\|^2 = \inf_{\substack{(c',\mathcal{U}')\in \mathrm{Alt}(c):\\ d_{\mathrm{H}}(c',c)=1}} \sum_{m=1}^{M} \lambda_m \|\mu(c_m) - \mu'(c'_m)\|^2.$$

- Instead of considering all the alternative instances in $\mathrm{Alt}(c)$, Lemma 2 shows it suffices to consider the instances whose partitions have a Hamming distance of 1 from the given partition c.
- Sketch of the proof: for any instance $(c^\dagger, \mathcal{U}^\dagger) \in \mathrm{Alt}(c)$ such that $d_{\mathrm{H}}(c^\dagger, c) > 1$, there exists another instance $(c^*, \mathcal{U}^*) \in \mathrm{Alt}(c)$ such that $d_{\mathrm{H}}(c^*, c) = 1$ and the objective function under $(c^*, \mathcal{U}^*)$ is not larger than that under $(c^\dagger, \mathcal{U}^\dagger)$.

Solution to Problem (InnerInf)

Proposition 1

For any $\lambda \in \mathcal{P}_M$ and $(c, \mathcal{U})$,

$$\inf_{(c', \mathcal{U}') \in \mathrm{Alt}(c)} \sum_{m=1}^{M} \lambda_m \|\mu(c_m) - \mu'(c'_m)\|^2$$

$$= \begin{cases} \min\limits_{\substack{k,k' \in [K]: \\ n(k) > 1, k' \neq k}} \frac{\bar{w}(k) w(k')}{\bar{w}(k) + w(k')} \|\mu(k) - \mu(k')\|^2 & \text{if } \lambda \in \mathcal{P}_M^+ \\ 0 & \text{otherwise} \end{cases}$$

where $w(k) := \sum_{m=1}^{M} \lambda_m \mathbb{1}\{c_m = k\}$, $n(k) := \sum_{m=1}^{M} \mathbb{1}\{c_m = k\}$ *and* $\bar{w}(k) := \min_{m \in [M]: c_m = k} \lambda_m$.

Solution to Problem (InnerInf)

Proposition 1

For any $\lambda \in \mathcal{P}_M$ and $(c, \mathcal{U})$,

$$\inf_{(c', \mathcal{U}') \in \mathrm{Alt}(c)} \sum_{m=1}^{M} \lambda_m \|\mu(c_m) - \mu'(c'_m)\|^2 = \begin{cases} \min\limits_{\substack{k,k' \in [K]: \\ n(k)>1, k' \neq k}} \frac{\bar{w}(k) w(k')}{\bar{w}(k) + w(k')} \|\mu(k) - \mu(k')\|^2 & \text{if } \lambda \in \mathcal{P}_M^+ \\ 0 & \text{otherwise} \end{cases}$$

where $w(k) := \sum_{m=1}^{M} \lambda_m \mathbb{1}\{c_m = k\}$, $n(k) := \sum_{m=1}^{M} \mathbb{1}\{c_m = k\}$ and $\bar{w}(k) := \min_{m \in [M]: c_m = k} \lambda_m$.

InnerInf becomes a finite minimization problem.

Continuity of Problem (InnerInf)

Proposition 2

For any fixed c, define $g : \mathcal{P}_M \times \mathbb{R}^{d\times K} \to \mathbb{R}^+$ as

$$g(\lambda,\mathcal{U}) := \inf_{(c',\mathcal{U}')\in \mathrm{Alt}(c)} \sum_{m=1}^{M} \lambda_m \|\mu(c_m) - \mu'(c'_m)\|^2.$$

Then g is continuous on $\mathcal{P}_M \times \mathscr{U}$.

Continuity of Problem (InnerInf)

Proposition 2

For any fixed c, define $g : \mathcal{P}_M \times \mathbb{R}^{d \times K} \to \mathbb{R}^+$ as

$$g(\lambda, \mathcal{U}) := \inf_{(c', \mathcal{U}') \in \mathrm{Alt}(c)} \sum_{m=1}^{M} \lambda_m \|\mu(c_m) - \mu'(c'_m)\|^2.$$

Then g is continuous on $\mathcal{P}_M \times \mathscr{U}$.

- Helps to assert that the stopping rule in our algorithm BOC is asymptotically optimal.

Solution to Problem (SupInf)

Recall that Problem (SupInf) is

$$D^*(c,\mathcal{U}) := \left(\frac{1}{2} \sup_{\lambda \in \mathcal{P}_M} \inf_{(c',\mathcal{U}') \in \mathrm{Alt}(c)} \sum_{m=1}^{M} \lambda_m \|\mu(c_m) - \mu'(c'_m)\|^2\right)^{-1}$$

Solution to Problem (SupInf)

Recall that Problem (SupInf) is

$$D^*(c,\mathcal{U}) := \left(\frac{1}{2}\sup_{\lambda\in\mathcal{P}_M}\inf_{(c',\mathcal{U}')\in\mathrm{Alt}(c)}\sum_{m=1}^{M}\lambda_m\|\mu(c_m)-\mu'(c'_m)\|^2\right)^{-1}$$

Proposition 3

For any $(c,\mathcal{U})$, $D^*(c,\mathcal{U})$ can be simplified as

$$D^*(c,\mathcal{U}) = 2\min_{w\in\mathcal{P}_K^+}\max_{\substack{k,k'\in[K]:\\ n(k)>1,k'\neq k}}\left(\frac{n(k)}{w(k)}+\frac{1}{w(k')}\right)\|\mu(k)-\mu(k')\|^{-2}.$$

Solution to Problem (SupInf)

Recall that Problem (SupInf) is

$$D^*(c,\mathcal{U}) := \left(\frac{1}{2} \sup_{\lambda \in \mathcal{P}_M} \inf_{(c',\mathcal{U}') \in \mathrm{Alt}(c)} \sum_{m=1}^{M} \lambda_m \|\mu(c_m) - \mu'(c'_m)\|^2 \right)^{-1}$$

Proposition 3

For any $(c,\mathcal{U})$, $D^*(c,\mathcal{U})$ can be simplified as

$$D^*(c,\mathcal{U}) = 2 \min_{w \in \mathcal{P}_K^+} \max_{\substack{k,k' \in [K]: \\ n(k) > 1, k' \neq k}} \left(\frac{n(k)}{w(k)} + \frac{1}{w(k')} \right) \|\mu(k) - \mu(k')\|^{-2}.$$

- Outer supremum in Problem (SupInf) is attained.

Solution to Problem (SupInf)

Proposition 4

For any $(c,\mathcal{U})$*, the solution to* $D^*(c,\mathcal{U})$

$$\underset{\lambda\in\mathcal{P}_M}{\arg\max} \inf_{(c',\mathcal{U}')\in\mathrm{Alt}(c)} \sum_{m=1}^{M} \lambda_m \|\mu(c_m) - \mu'(c'_m)\|^2 \tag{1}$$

is unique. If λ^* *denotes the solution to* (1) *and* w^* *denotes the solution to*

$$\underset{w\in\mathcal{P}_K^+}{\arg\min} \max_{\substack{k,k'\in[K]:\\ n(k)>1,k'\neq k}} \left(\frac{n(k)}{w(k)} + \frac{1}{w(k')}\right) \|\mu(k) - \mu(k')\|^{-2},$$

then λ^* *can be expressed in terms of* w^* *as (bijection between* λ^* *and* w^**)*

$$\lambda_m^* = \frac{w^*(c_m)}{n(c_m)} \text{ for all } m \in [M].$$

Continuity of Problem (SupInf)

Proposition 5

For any fixed c, define $\Lambda : \mathbb{R}^{d\times K} \to \mathcal{P}_M$ as

$$\Lambda(\mathcal{U}) := \arg\max_{\lambda\in\mathcal{P}_M} \inf_{(c',\mathcal{U}')\in\mathrm{Alt}(c)} \sum_{m=1}^{M} \lambda_m \|\mu(c_m) - \mu'(c'_m)\|^2$$

where $\mathcal{U} = [\mu(1), \mu(2), \ldots, \mu(K)] \in \mathbb{R}^{d\times K}$. Then Λ is continuous on $\mathscr{U}$.

Continuity of Problem (SupInf)

Proposition 5

For any fixed c, define $\Lambda : \mathbb{R}^{d\times K} \to \mathcal{P}_M$ as

$$\Lambda(\mathcal{U}) := \operatorname*{arg\,max}_{\lambda\in\mathcal{P}_M} \inf_{(c',\mathcal{U}')\in\mathrm{Alt}(c)} \sum_{m=1}^{M} \lambda_m \|\mu(c_m) - \mu'(c'_m)\|^2$$

where $\mathcal{U} = [\mu(1), \mu(2), \ldots, \mu(K)] \in \mathbb{R}^{d\times K}$. Then Λ is continuous on $\mathscr{U}$.

- The correspondence $\Lambda(\mathcal{U})$ is single-valued and upper hemicontinuous.
- A single-valued correspondence that is hemicontinuous is continuous (Sundaram, 1996).
- Guarantees computationally efficiency and asymptotic optimality of our sampling rule.

Two Optimization Problems

Simplifications of InnerInf and SupInf

- Problem (InnerInf) $\Leftrightarrow$ A finite minimization problem
- Problem (SupInf) $\Leftrightarrow$ A finite convex minimax problem (Gigola and Gomez, 1990; Herrmann, 1999; Gaudioso et al., 2006)

Implications of simplifications

- Problem (InnerInf) plays an essential role in the computation of the stopping rule of our method.
- Problem (SupInf) guarantees the computationally efficiency and the asymptotic optimality of our sampling rule.

Outline

Weighted K-means with Maximin Initialization

Our Goal at Each Step

- Although we only aim at producing a correct partition in the final recommendation rule, learning the K unknown mean vectors of the clusters is essential in the sampling rule as well as the stopping rule.

Weighted K-means with Maximin Initialization

Our Goal at Each Step

- Although we only aim at producing a correct partition in the final recommendation rule, learning the K unknown mean vectors of the clusters is essential in the sampling rule as well as the stopping rule.
- Question: Given some past measurements on the arm set, how to produce an estimate of the pair $(c, \mathcal{U})$?

Weighted K-means with Maximin Initialization

- Given the past arm pulls and observations up to time t, the log-likelihood function that the instance is $(c', \mathcal{U}')$ is

$$\ell(c', \mathcal{U}' \mid A_1, X_1, \ldots, A_t, X_t) := -\frac{1}{2}\sum_{\bar{t}=1}^{t} \|X_{\bar{t}} - \mu'(c'_{A_{\bar{t}}})\|^2 - \frac{td}{2}\log(2\pi).$$

Weighted K-means with Maximin Initialization

- Given the past arm pulls and observations up to time t, the log-likelihood function that the instance is $(c', \mathcal{U}')$ is

$$\ell(c', \mathcal{U}' \mid A_1, X_1, \ldots, A_t, X_t) := -\frac{1}{2} \sum_{\bar{t}=1}^{t} \|X_{\bar{t}} - \mu'(c'_{A_{\bar{t}}})\|^2 - \frac{td}{2} \log(2\pi).$$

- For any arm $m \in [M]$, let $N_m(t)$ and $\hat{\mu}_m(t)$ denote the number of pulls and the empirical estimate up to time t, respectively.

Weighted K-means with Maximin Initialization

- Given the past arm pulls and observations up to time t, the log-likelihood function that the instance is $(c', \mathcal{U}')$ is

$$\ell(c', \mathcal{U}' \mid A_1, X_1, \ldots, A_t, X_t) := -\frac{1}{2}\sum_{\bar{t}=1}^{t} \|X_{\bar{t}} - \mu'(c'_{A_{\bar{t}}})\|^2 - \frac{td}{2}\log(2\pi).$$

- For any arm $m \in [M]$, let $N_m(t)$ and $\hat{\mu}_m(t)$ denote the number of pulls and the empirical estimate up to time t, respectively.
- The maximum likelihood estimate of the unknown pair $(c, \mathcal{U})$ is

$$\underset{(c', \mathcal{U}')}{\arg\min} \sum_{m=1}^{M} N_m(t) \|\hat{\mu}_m(t) - \mu'(c'_m)\|^2$$

which involves minimizing a weighted sum of squared distances between the empirical estimate of each arm and its associated center.

Weighted K-means with Maximin Initialization

Weighted K-means problem

$$\arg\min_{(c',\mathcal{U}')} \sum_{m=1}^{M} N_m(t)\|\hat{\mu}_m(t) - \mu'(c'_m)\|^2$$

Weighted K-means with Maximin Initialization

Weighted K-means problem

$$\arg\min_{(c',\mathcal{U}')} \sum_{m=1}^{M} N_m(t)\|\hat{\mu}_m(t) - \mu'(c'_m)\|^2$$

Remarks

- MLE $\Leftrightarrow$ The classical offline weighted K-means clustering problem

Weighted K-means with Maximin Initialization

Weighted K-means problem

$$\arg\min_{(c',\mathcal{U}')} \sum_{m=1}^{M} N_m(t)\|\hat{\mu}_m(t) - \mu'(c'_m)\|^2$$

Remarks

- MLE $\Leftrightarrow$ The classical offline weighted K-means clustering problem
- Any algorithm designed for the weighted K-means clustering problem is applicable to obtain an approximate (not exact) solution to the MLE problem.

Weighted K-means with Maximin Initialization

Weighted K-means problem

$$\arg\min_{(c',\mathcal{U}')} \sum_{m=1}^{M} N_m(t)\|\hat{\mu}_m(t) - \mu'(c'_m)\|^2$$

Remarks

- MLE $\Leftrightarrow$ The classical offline weighted K-means clustering problem
- Any algorithm designed for the weighted K-means clustering problem is applicable to obtain an approximate (not exact) solution to the MLE problem.
- No theoretical guarantees for finding a global minimum of this problem in general.

Weighted K-means with Maximin Initialization

Weighted K-means problem

$$\arg\min_{(c',\mathcal{U}')} \sum_{m=1}^{M} N_m(t) \|\hat{\mu}_m(t) - \mu'(c'_m)\|^2$$

Remarks

- MLE $\Leftrightarrow$ The classical offline weighted K-means clustering problem
- Any algorithm designed for the weighted K-means clustering problem is applicable to obtain an approximate (not exact) solution to the MLE problem.
- No theoretical guarantees for finding a global minimum of this problem in general.
- Weighted K-means with Maximin Initialization (Gonzalez, 1985) possesses useful properties!

Weighted K-means with Maximin Initialization

Algorithm 1 Weighted K-means with Maximin Initialization (K-means–Maximin)

Input: Number of clusters K, empirical estimate $\hat{\mu}_m$ and weighting N_m for all $m \in [M]$

1: Choose the empirical estimate of an arbitrary arm as the first cluster center $\hat{\mu}(1)$

2: **for** $k = 2$ **to** K **do** ▷ Maximin Initialization

3: Choose the empirical estimate of the arm that has the greatest Euclidean distance to the nearest existing center as the k-th center $\hat{\mu}(k)$:

$$\hat{\mu}(k) = \arg\max_{m \in [M]} \min_{1 \le k' \le k-1} \|\hat{\mu}_m - \hat{\mu}(k')\|$$

4: **end for**

Weighted K-means with Maximin Initialization

repeat ▷ Weighted K-means

Assign each arm to its closest cluster center:

$$\hat{c}_m = \arg\min_{k \in [K]} \|\hat{\mu}_m - \hat{\mu}(k)\|$$

Update each cluster center as the weighted mean of the empirical estimates of the arms in it:

$$\hat{\mu}(k) = \frac{\sum_{m \in [M]} N_m \hat{\mu}_m \mathbb{1}\{\hat{c}_m = k\}}{\sum_{m \in [M]} N_m \mathbb{1}\{\hat{c}_m = k\}}$$

until Clustering $\hat{c}$ no longer changes

Set $\mu^{\mathrm{out}}(k) = \hat{\mu}(k)$ for all $k \in [K]$

Output: $c^{\mathrm{out}} = \hat{c}$ and $\mathcal{U}^{\mathrm{out}} = [\mu^{\mathrm{out}}(1), \mu^{\mathrm{out}}(2), \ldots, \mu^{\mathrm{out}}(K)]$

Stopping Rule

Usual Strategy

- As the arm sampling proceeds, the algorithm needs to determine when to stop the sampling and then to recommend a partition with an error probability of at most δ.

Stopping Rule

Usual Strategy

- As the arm sampling proceeds, the algorithm needs to determine when to stop the sampling and then to recommend a partition with an error probability of at most δ.
- Most existing algorithms for pure exploration in the fixed-confidence setting (e.g., Garivier and Kaufmann (2016), Jedra and Proutiere (2020), Feng et al. (2021), Réda et al. (2021)) consider the Generalized Likelihood Ratio (GLR) statistic and find suitable task-specific threshold functions.

Stopping Rule: Problems with the GLR

- Method based on the GLR is computationally intractable.

Stopping Rule: Problems with the GLR

- Method based on the GLR is computationally intractable.
- Let $(c^{t*}, \mathcal{U}^{t*})$ be the maximum likelihood estimate at time t.

Stopping Rule: Problems with the GLR

- Method based on the GLR is computationally intractable.
- Let $(c^{t*}, \mathcal{U}^{t*})$ be the maximum likelihood estimate at time t.
- The logarithm of the GLR statistic for testing $(c^{t*}, \mathcal{U}^{t*})$ against its alternative instances can be written as

$$\log\text{-GLR}(c^{t*}, \mathcal{U}^{t*}) = \frac{1}{2}\Big(- \sum_{m=1}^{M} N_m(t) \|\hat{\mu}_m(t) - \mu^{t*}(c_m^{t*})\|^2 + \min_{(c', \mathcal{U}') \in \mathrm{Alt}(c^{t*})} \sum_{m=1}^{M} N_m(t) \|\hat{\mu}_m(t) - \mu'(c'_m)\|^2 \Big).$$

Stopping Rule: Problems with the GLR

- Method based on the GLR is computationally intractable.
- Let $(c^{t*}, \mathcal{U}^{t*})$ be the maximum likelihood estimate at time t.
- The logarithm of the GLR statistic for testing $(c^{t*}, \mathcal{U}^{t*})$ against its alternative instances can be written as

$$\log\text{-GLR}(c^{t*}, \mathcal{U}^{t*}) = \frac{1}{2}\Bigg(-\sum_{m=1}^{M} N_m(t) \|\hat{\mu}_m(t) - \mu^{t*}(c_m^{t*})\|^2 + \min_{(c', \mathcal{U}') \in \mathrm{Alt}(c^{t*})} \sum_{m=1}^{M} N_m(t) \|\hat{\mu}_m(t) - \mu'(c'_m)\|^2 \Bigg).$$

- Problem 1: Requires exact global minimizer in the log-GLR statistic.

Stopping Rule: Problems with the GLR

- Method based on the GLR is computationally intractable.
- Let $(c^{t*}, \mathcal{U}^{t*})$ be the maximum likelihood estimate at time t.
- The logarithm of the GLR statistic for testing $(c^{t*}, \mathcal{U}^{t*})$ against its alternative instances can be written as

$$\log\text{-GLR}(c^{t*}, \mathcal{U}^{t*}) = \frac{1}{2}\Big(-\sum_{m=1}^{M} N_m(t)\|\hat{\mu}_m(t) - \mu^{t*}(c_m^{t*})\|^2 + \min_{(c', \mathcal{U}') \in \mathrm{Alt}(c^{t*})} \sum_{m=1}^{M} N_m(t)\|\hat{\mu}_m(t) - \mu'(c'_m)\|^2 \Big).$$

- Problem 1: Requires exact global minimizer in the log-GLR statistic.
- Problem 2: Similar Hamming distance 1 nice property (Lemma 2) does not hold when $\mu(c_m)$ is replaced by $\hat{\mu}_m(t)$ $\Longrightarrow$ Mismatch!

Stopping Rule: Our Remedy

- Let $(c^{t-1}, \mathcal{U}^{t-1})$ be the estimate of the true pair $(c, \mathcal{U})$ produced by the K-means–Maximin based on the past measurements up to time $t-1$.

Stopping Rule: Our Remedy

- Let $(c^{t-1}, \mathcal{U}^{t-1})$ be the estimate of the true pair $(c, \mathcal{U})$ produced by the K-MEANS–MAXIMIN based on the past measurements up to time $t-1$.
- Consider the statistic

$$Z(t) := \frac{1}{2}\left(\left(-\sqrt{Z_1(t)} + \sqrt{Z_2(t)}\right)_+\right)^2$$

with

$$Z_1(t) := \sum_{m=1}^{M} N_m(t) \|\hat{\mu}_m(t) - \mu^{t-1}(c_m^{t-1})\|^2$$

Stopping Rule: Our Remedy

- Let $(c^{t-1}, \mathcal{U}^{t-1})$ be the estimate of the true pair $(c, \mathcal{U})$ produced by the K-means–Maximin based on the past measurements up to time $t-1$.
- Consider the statistic

$$Z(t) := \frac{1}{2}\left(\left(-\sqrt{Z_1(t)} + \sqrt{Z_2(t)}\right)_+\right)^2$$

with

$$Z_1(t) := \sum_{m=1}^{M} N_m(t) \|\hat{\mu}_m(t) - \mu^{t-1}(c_m^{t-1})\|^2$$

and

$$Z_2(t) := \min_{(c', \mathcal{U}') \in \mathrm{Alt}(c^{t-1})} \sum_{m=1}^{M} N_m(t) \|\mu^{t-1}(c_m^{t-1}) - \mu'(c'_m)\|^2,$$

which is analogous to

Problem (InnerInf): $$\inf_{(c', \mathcal{U}') \in \mathrm{Alt}(c)} \sum_{m=1}^{M} \lambda_m \|\mu(c_m) - \mu'(c'_m)\|^2$$

Stopping Rule: Our Remedy

- Let $(c^{t-1}, \mathcal{U}^{t-1})$ be the estimate of the true pair $(c, \mathcal{U})$ produced by the K-MEANS–MAXIMIN based on the past measurements up to time $t-1$.
- Consider the statistic

$$Z(t) := \frac{1}{2}\left(\left(-\sqrt{Z_1(t)} + \sqrt{Z_2(t)}\right)_+\right)^2$$

with

$$Z_1(t) := \sum_{m=1}^{M} N_m(t)\|\hat{\mu}_m(t) - \mu^{t-1}(c_m^{t-1})\|^2$$

and

$$Z_2(t) := \min_{(c',\mathcal{U}')\in\mathrm{Alt}(c^{t-1})} \sum_{m=1}^{M} N_m(t)\|\mu^{t-1}(c_m^{t-1}) - \mu'(c'_m)\|^2,$$

which is analogous to

$$\text{Problem (InnerInf):} \quad \inf_{(c',\mathcal{U}')\in\mathrm{Alt}(c)} \sum_{m=1}^{M} \lambda_m\|\mu(c_m) - \mu'(c'_m)\|^2 \quad \text{Match!}$$

Stopping Rule

Stopping Time

The stopping time is defined as

$$\tau_\delta := \inf\{t \in \mathbb{N} : Z(t) \geq \beta(\delta, t)\}$$

where $\beta(\delta, t)$ is a threshold function inspired by the concentration results for univariate Gaussian distributions (Kaufmann and Koolen, 2021).

Stopping Rule

Stopping Time

The stopping time is defined as

$$\tau_\delta := \inf\{t \in \mathbb{N} : Z(t) \geq \beta(\delta, t)\}$$

where $\beta(\delta, t)$ is a threshold function inspired by the concentration results for univariate Gaussian distributions (Kaufmann and Koolen, 2021).

Proposition 6

The stopping time satisfies that

$$\Pr(\tau_\delta < \infty, c^{\mathrm{out}} \not\sim c) \leq \delta.$$

Algorithm: Bandit Online Clustering (BOC)

Algorithm 2 Bandit Online Clustering (BOC)

Input: Number of clusters K, confidence level δ and arm set $[M]$

Sample each arm once, set $t = M$ and initialize $\hat{\mu}_m(t)$ and $N_m(t) = 1$ for all $m \in [M]$.
repeat
  if $\min_{m\in[M]} N_m(t) \le \max(\sqrt{t} - M/2, 0)$ **then** ▷ Forced exploration
    Sample $A_{t+1} = \arg\min_{m\in[M]} N_m(t)$ and $(c^t, \mathcal{U}^t) \leftarrow (c^{t-1}, \mathcal{U}^{t-1})$
  else
    $(c^t, \mathcal{U}^t) \leftarrow \textsc{K-means–Maximin}(K, \{\hat{\mu}_m(t)\}_{m\in[M]}, \{N_m(t)\}_{m\in[M]})$
    Solve ▷ Problem (SupInf)

$$\lambda^*(t) = \arg\max_{\lambda\in\mathcal{P}_M} \inf_{(c',\mathcal{U}')\in \mathrm{Alt}(c^t)} \sum_{m=1}^{M} \lambda_m \left\|\mu^t(c_m^t) - \mu'(c'_m)\right\|^2$$

    Sample $A_{t+1} = \arg\max_{m\in[M]}(t\lambda_m^*(t) - N_m(t))$
  end if
  $t \leftarrow t + 1$, update $\hat{\mu}_m(t)$ and $N_m(t)$ for all $m \in [M]$

Algorithm

11: Compute

$$Z_1(t) = \sum_{m=1}^{M} N_m(t) \|\hat{\mu}_m(t) - \mu^{t-1}(c_m^{t-1})\|^2$$

and solve ▷ Problem (InnerInf)

$$Z_2(t) = \min_{(c',\mathcal{U}') \in \mathrm{Alt}(c^{t-1})} \sum_{m=1}^{M} N_m(t) \|\mu^{t-1}(c_m^{t-1}) - \mu'(c'_m)\|^2$$

12: Set

$$Z(t) = \frac{1}{2} \left(\left(-\sqrt{Z_1(t)} + \sqrt{Z_2(t)} \right)_+ \right)^2$$

13: **until** $Z(t) \geq \beta(\delta, t)$

Output: $\tau_\delta = t$ and $c^{\mathrm{out}} = c^{t-1}$

Sample Complexity

Theorem 3

For any instance $(c, \mathcal{U})$, *Bandit Online Clustering ensures that* $\Pr(\tau_\delta < \infty) = 1$ *and*

$$\limsup_{\delta \to 0} \frac{\mathbb{E}[\tau_\delta]}{\log(1/\delta)} \leq D^*(c, \mathcal{U}).$$

Hence, combining this with the lower bound,

$$\lim_{\delta \to 0} \frac{\mathbb{E}[\tau_\delta^*]}{\log(1/\delta)} = D^*(c, \mathcal{U}).$$

Sample Complexity

Theorem 3

For any instance $(c, \mathcal{U})$*, Bandit Online Clustering ensures that* $\Pr(\tau_\delta < \infty) = 1$ *and*

$$\limsup_{\delta \to 0} \frac{\mathbb{E}[\tau_\delta]}{\log(1/\delta)} \le D^*(c, \mathcal{U}).$$

Hence, combining this with the lower bound,

$$\lim_{\delta \to 0} \frac{\mathbb{E}[\tau^*_\delta]}{\log(1/\delta)} = D^*(c, \mathcal{U}).$$

- The expected sample complexity of BOC asymptotically matches the instance-dependent lower bound as the confidence level $\delta \to 0$.

Sample Complexity

Theorem 3

For any instance $(c, \mathcal{U})$*, Bandit Online Clustering ensures that* $\Pr(\tau_\delta < \infty) = 1$ *and*

$$\limsup_{\delta \to 0} \frac{\mathbb{E}[\tau_\delta]}{\log(1/\delta)} \leq D^*(c, \mathcal{U}).$$

Hence, combining this with the lower bound,

$$\lim_{\delta \to 0} \frac{\mathbb{E}[\tau_\delta^*]}{\log(1/\delta)} = D^*(c, \mathcal{U}).$$

- The expected sample complexity of BOC asymptotically matches the instance-dependent lower bound as the confidence level $\delta \to 0$.
- It is also computationally efficient in terms of its sampling, stopping and recommendation rules.

Outline

Synthetic Dataset: Verifying the Asymptotic Behavior

- Three synthetic instances with varying difficulty levels, where $K = 4$, $M = 11$ and $d = 3$.
- The partitions and the first three cluster centers of all the three instances are the same, while their fourth cluster centers vary.

Synthetic Dataset: Verifying the Asymptotic Behavior

- Three synthetic instances with varying difficulty levels, where $K = 4$, $M = 11$ and $d = 3$.
- The partitions and the first three cluster centers of all the three instances are the same, while their fourth cluster centers vary.
- $$\begin{cases} c = [1, 1, 2, 2, 2, 2, 3, 3, 3, 4, 4] \\ \mu(1) = [0, 0, 0]^\top \\ \mu(2) = [0, 10, 0]^\top \\ \mu(3) = [0, 0, 10]^\top \\ \mu(4) = \begin{cases} [5, 0, 0]^\top & \text{for the \textit{easy} instance,} \\ [1, 0, 0]^\top & \text{for the \textit{moderate} instance,} \\ [0.5, 0, 0]^\top & \text{for the \textit{challenging} instance} \end{cases} \end{cases}$$

Synthetic Dataset: Verifying the Asymptotic Behavior

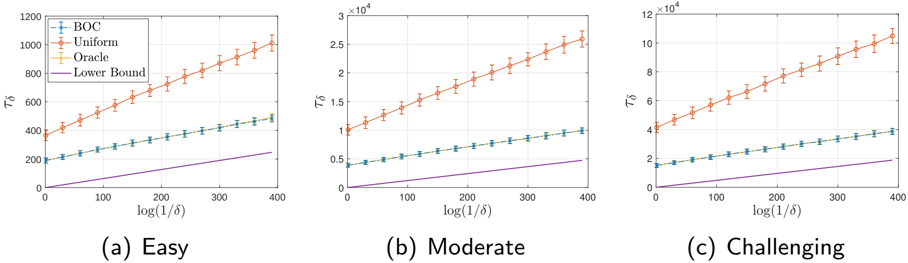


Figure: The empirical averaged sample complexities of the different methods (BOC, Uniform, Oracle) with respect to $\log(1/\delta)$.

Thanks for listening!

References

Herman Chernoff. Sequential design of experiments. *The Annals of Mathematical Statistics*, 30(3):755–770, 1959.

Tze Leung Lai and Herbert Robbins. Asymptotically efficient adaptive allocation rules. *Advances in Applied Mathematics*, 6(1): 4–22, 1985.

Tor Lattimore and Csaba Szepesvári. The end of optimism? An asymptotic analysis of finite-armed linear bandits. In *International Conference on Artificial Intelligence and Statistics (AISTATS)*, pages 728–737. PMLR, 2017.

Emilie Kaufmann, Olivier Cappé, and Aurélien Garivier. On the complexity of best-arm identification in multi-armed bandit models. *Journal of Machine Learning Research*, 17(1):1–42, 2016.

Nidhin Koshy Vaidhiyan and Rajesh Sundaresan. Learning to detect an oddball target. *IEEE Transactions on Information Theory*, 64(2):831–852, 2017.

Rangarajan K Sundaram. *A First Course in Optimization Theory*. Cambridge University Press, 1996.

Cristina Gigola and Susana Gomez. A regularization method for solving the finite convex min-max problem. *SIAM Journal on Numerical Analysis*, 27(6):1621–1634, 1990.

Jeffrey W Herrmann. A genetic algorithm for minimax optimization problems. In *Proceedings of the 1999 Congress on Evolutionary Computation-CEC99 (Cat. No. 99TH8406)*, volume 2, pages 1099–1103. IEEE, 1999.

Manlio Gaudioso, Giovanni Giallombardo, and Giovanna Miglionico. An incremental method for solving convex finite min-max problems. *Mathematics of Operations Research*, 31(1):173–187, 2006.

Teofilo F Gonzalez. Clustering to minimize the maximum intercluster distance. *Theoretical Computer Science*, 38:293–306, 1985.

Aurélien Garivier and Emilie Kaufmann. Optimal best arm identification with fixed confidence. In *Conference on Learning Theory*, pages 998–1027. PMLR, 2016.

Yassir Jedra and Alexandre Proutiere. Optimal best-arm identification in linear bandits. *Advances in Neural Information Processing Systems*, 33:10007–10017, 2020.

Yifan Feng, Rene Caldentey, and Christopher Thomas Ryan. Robust learning of consumer preferences. *Operations Research*, 2021.

Clémence Réda, Andrea Tirinzoni, and Rémy Degenne. Dealing with misspecification in fixed-confidence linear top-m identification. *Advances in Neural Information Processing Systems*, 34, 2021.

Emilie Kaufmann and Wouter M Koolen. Mixture martingales revisited with applications to sequential tests and confidence intervals. *Journal of Machine Learning Research*, 22(246):1–44, 2021.