The price of unfairness in linear bandits with biased feedback

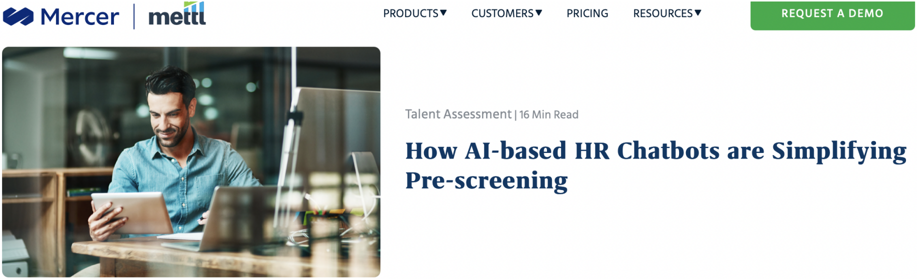
Solenne Gaucher(1), Alexandra Carpentier(2) et Christophe Giraud(1)

(1) Université Paris Saclay
(2) Postdam Universität

London, June 2022

Fairness in Machine Learning: a major societal concern

Machine Learning is ubiquitous in daily life



Fairness in Machine Learning: a major societal concern

Machine Learning is ubiquitous in daily life

05-17-19

Schools are using software to help pick who gets in. What could go wrong?

Admissions officers are increasingly turning to automation and AI with the hope of streamlining the application process and leveling the playing field.

Fairness in Machine Learning: a major societal concern

Machine Learning is ubiquitous in daily life

SCIENCE ADVANCES | RESEARCH ARTICLE

RESEARCH METHODS

The accuracy, fairness, and limits of predicting recidivism

Julia Dressel and Hany Farid*

Algorithms for predicting recidivism are commonly used to assess a criminal defendant's likelihood of committing a crime. These predictions are used in pretrial, parole, and sentencing decisions. Proponents of these systems argue that big data and advanced machine learning make these analyses more accurate and less biased than humans. We show, however, that the widely used commercial risk assessment software COMPAS is no more accurate or fair than predictions made by people with little or no criminal justice expertise. In addition, despite COMPAS's collection of 137 features, the same accuracy can be achieved with a simple linear predictor with only two features.

A simple fairness problem in sequential decision making

Sequential decision making with covariates

Actions and rewards

Actions: indexed by $\mathcal{X} \subset \mathbb{R}^d$
Reward: $f(x)$ for action $x \in \mathcal{X}$ for some unobserved $f : \mathcal{X} \to \mathbb{R}$.

Bandit problems with covariates

At each round $t = 1, \ldots, T$

- the agent chooses an action $x_t \in \mathcal{X}$, based on her historical data
- she observes the feedback $y_t = f(x_t) + \xi_t$, with $(\xi_t)_{t \geq 1}$ independent.

Goal

Maximise the unobserved cumulated reward $\sum_t f(x_t)$.

Sequential decision making with covariates

Optimal oracle strategy

Sample at each round $x^* \in \underset{x \in \mathcal{X}}{\operatorname{argmax}} f(x)$.

Infeasible since x^* is unknown...

Decision maker objective

Minimize the regret

$$R_T = \mathbb{E}\Big[\sum_{t \leq T} (f(x^*) - f(x_t))\Big]$$

Linear bandit: $f(x) = x^\top \theta^*$ with θ^* unknown (very popular in applications)

Fairness issue

Unfairness in ML

The main cause of unfairness in applications of ML algorithms, is the presence of biases in the data.

Setting

- Each action x is characterized by an (observed) attribute $z_x \in \{-1, +1\}$ (e.g. gender)
- The feedbacks are biased depending on z_x

Questions:

1. What is the impact of such a bias?
2. How do handle it?

Biased Linear Bandits

Linear bandit with biased feedbacks

At each round $t = 1, \dots, T$

- the agent chooses an action $x_t \in \mathcal{X} \subset \mathbb{R}^d$, whose sensitive attribute is $z_{x_t} \in \{-1, 1\}$
- she receives the unobserved reward $x_t^\top \gamma^*$;
- she observes the biased feedback $y_t = x_t^\top \gamma^* + z_{x_t}\omega^* + \xi_t$.

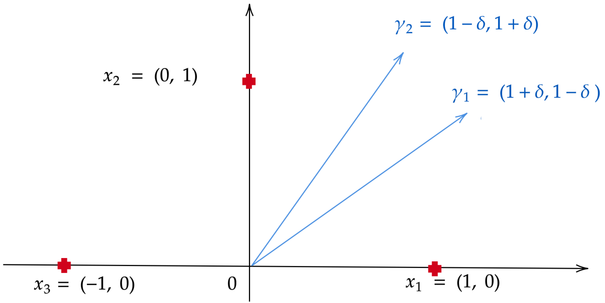
Objective

Minimize the regret

$$R_T = \mathbb{E}\Big[\sum_{t \le T} (x^* - x_t)^\top \gamma^*\Big], \quad \text{where} \quad x^* \in \operatorname*{argmax}_{x \in \mathcal{X}} x^\top \gamma^*.$$

An insightful toy example

Toy example - unbiased feedbacks

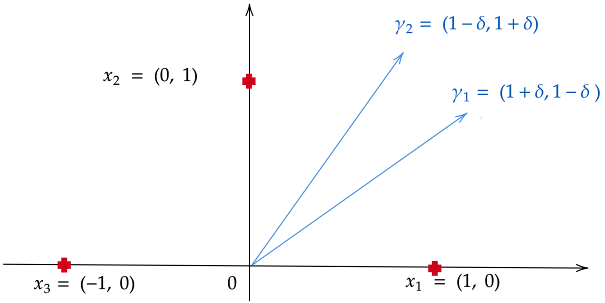


Unbiased feedback: $\gamma^T x + \xi$ with ξ sub-Gaussian

Best action:

- if $\gamma = \gamma_k$ action x_k is optimal, $k = 1, 2$;
- in both cases x_3 is very suboptimal.

Toy example - unbiased feedbacks



1. The feedbacks differ by 2δ between x_1 and x_2
2. Confidence intervals have width $\propto \sqrt{\log(\text{confidence}^{-1})/N_{x_k}(t)}$

So:

- if $N_{x_k}(T) \lesssim \delta^{-2}$, for $k = 1$ or 2: we cannot find the best action, and $R_T = \Theta(T\delta)$;
- if $N_{x_k}(T) \gtrsim \delta^{-2} \log(T)$, for $k = 1$ and 2: we find the best action with confidence $1/T$, and $R_T = \Theta(\delta \cdot \delta^{-2} \log(T)) = \Theta(\delta^{-1} \log(T))$;

Toy example - unbiased feedbacks

- if $N_{x_k}(T) \leq \delta^{-2}$, for $k = 1$ or 2: we cannot find the best action, and $R_T = \Theta(T\delta)$;
- if $N_{x_k}(T) \geq \delta^{-2}\log(T)$, for $k = 1$ and 2: we find the best action with confidence $1/T$, and the regret is $R_T = \Theta(\delta^{-1}\log(T))$;

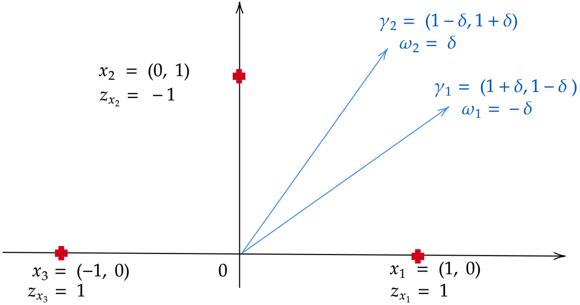
Optimal regret with unbiased feedbacks

Large T regret: $R_T = \Theta(\delta^{-1}\log(T))$, when $T \to \infty$;

Worst case regret: the worst case is when $\delta^{-2} = T$, and then

$$R_T = \Theta(T\delta) = \Theta(\sqrt{T}).$$

Toy example - biased feedbacks



Biased feedback: $\gamma^T x + z_x \omega + \xi$ with ξ sub-Gaussian.

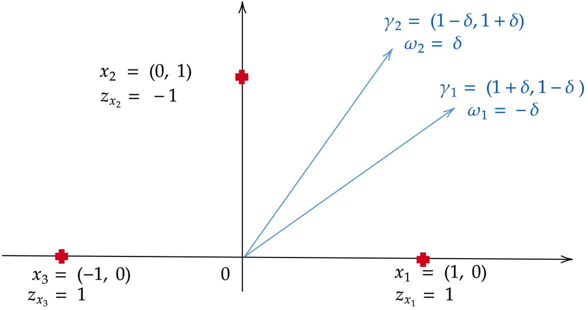
For (γ_1, ω_1) and (γ_2, ω_2), the feedbacks are identical for x_1, x_2:

$$x_1{}^\top \gamma_1 + z_{x_1}\omega_1 = x_1{}^\top \gamma_2 + z_{x_1}\omega_2 = 1,$$

$$\text{and} \quad x_2{}^\top \gamma_1 + z_{x_2}\omega_1 = x_2{}^\top \gamma_2 + z_{x_2}\omega_2 = 1.$$

$\Longrightarrow$ We need to sample the very sub-optimal action x_3 to discriminate between (γ_1, ω_1) and (γ_2, ω_2).

Toy example - biased feedbacks



The feedback when choosing x_3 differs by 4δ between (γ_1, ω_1) and (γ_2, ω_2):

- if $N_{x_3}(T) \lesssim \delta^{-2}$: we cannot find the best action, and $R_T = \Theta(T\delta)$;
- if $N_{x_3}(T) \gtrsim \delta^{-2}\log(T)$: we find the best and $R_T = \Theta(\delta^{-2}\log(T))$.

Worst-case regret is achieved for $\delta = T^{-1/3}$, and $R_T = \tilde{\Theta}(T^{2/3})$.

Toy example: summary

	Unbiased	Biased
Asymptotic regret	$R_T = \Theta(\delta^{-1}\log(T))$	$R_T = \Theta(\delta^{-2}\log(T))$
Worst case regret	for $\delta = T^{-1/2}$ $R_T^* = \Theta(\sqrt{T})$	for $\delta = T^{-1/3}$ $R_T^* = \tilde{\Theta}(T^{2/3})$

Question: What is the price of biased feedbacks in general?

Refresher on unbiased linear bandits

A general recipe in bandits: successive elimination

Confidence interval

$\mathcal{I}_x(t) :=$ confidence interval (at a prescribed level) for $f(x)$ from the data collected up to time t.

Recipe

If "$\mathcal{I}_x(t) < \max_{x'} \mathcal{I}_{x'}(t)$", drop out the action x from $\mathcal{X}$.

If "$\mathcal{I}_x(t) \cap \max_{x'} \mathcal{I}_{x'}(t) \neq \emptyset$" then get more samples to shrink $\mathcal{I}_x$

Successive Elimination (principle)

<u>REPEAT</u> for $\varepsilon \searrow 0$

- Sample a minimal number of actions from $\mathcal{X}$ to get

$$|\mathcal{I}_x(t)| \leq \varepsilon \quad \text{for all} \quad x \in \mathcal{X};$$

- Drop out from $\mathcal{X}$ all actions x such that $\mathcal{I}_x(t) < \max_{x'} \mathcal{I}_{x'}(t)$.

Linear bandits

Unbiased linear bandit

Reward: $f(x) = x^\top \theta^*$

Feedback: $y = x^\top \theta^* + \xi$ with ξ subGaussian(1).

Assumptions

$|\mathcal{X}| = k < \infty$ and $|x^\top \theta^*| \leq 1$ for all $x \in \mathcal{X}$.

Confidence bounds

OLS estimator

For n sampled actions $x_1, ..., x_n$ in $\mathcal{X}$, the OLS estimator is

$$\widehat{\theta} = V^+ \sum_{s \leq n} x_s y_s, \quad \text{where} \quad V = \sum_{s \leq n} x_s x_s^\top,$$

and V^+ is the Moore-Penrose pseudo inverse.

Confidence bound

If $x_1, ..., x_n$ are fixed, then for all $x \in \mathrm{Range}(V)$,

$$\mathbb{P}\left(\left| \left(\widehat{\theta} - \theta^*\right)^\top x \right| \leq \sqrt{2 \|x\|_{V^+}^2 \log\left(\frac{1}{\delta}\right)} \right) \geq 1 - \delta.$$

where $\|x\|_{V^+}^2 := x^\top V^+ x$.

Confidence bounds

OLS estimator

For n sampled actions $x_1, \ldots, x_n$ in $\mathcal{X}$, the OLS estimator is

$$\widehat{\theta} = V^+ \sum_{s \le n} x_s y_s, \quad \text{where} \quad V = \sum_{s \le n} x_s x_s^\top,$$

and V^+ is the Moore-Penrose pseudo inverse.

Confidence bound

If $x_1, \ldots, x_n$ are fixed, then for all $x \in \mathrm{Range}(V)$,

$$\mathbb{P}\left(\left| \left(\widehat{\theta} - \theta^* \right)^\top x \right| \le \sqrt{2 \, \|x\|_{V^+}^2 \log\left(\frac{1}{\delta}\right)} \right) \ge 1 - \delta.$$

where $\|x\|_{V^+}^2 := x^\top V^+ x$.

G-optimal design

If we choose each action x exactly $\mu(x)$ times

$$\sum_{s \le n} x_s x_s^\top = V(\mu) := \sum_{x \in \mathcal{X}} \mu(x) x x^\top$$

G-optimal design

$$\mu_n^* \in \operatorname*{argmin}_{|\mu|=n} \max_{x \in \mathcal{X}} \|x\|^2_{V(\mu)^+} . \qquad \text{(G-optimal design)}$$

fulfills

$$\max_{x \in \mathcal{X}} \|x\|^2_{V(\mu_n^*)^+} \le \frac{d}{n}.$$

Confidence bound

If each action $x \in \mathcal{X}$ is sampled $\mu_n^*(x)$ times with $n = \frac{2d}{\epsilon^2} \log\left(k\delta^{-1}\right)$,

then

$$\max_{x \in \mathcal{X}} \left| \left(\widehat{\theta} - \theta^*\right)^\top x \right| \le \epsilon, \quad \text{with probability at least } 1 - \delta.$$

G-optimal design

If we choose each action x exactly $\mu(x)$ times

$$\sum_{s\leq n} x_s x_s^\top = V(\mu) := \sum_{x\in\mathcal{X}} \mu(x) x x^\top$$

G-optimal design

$$\mu_n^* \in \underset{|\mu|=n}{\operatorname{argmin}} \ \max_{x\in\mathcal{X}} \|x\|^2_{V(\mu)^+} \ . \qquad \text{(G-optimal design)}$$

fulfills

$$\max_{x\in\mathcal{X}} \ \|x\|^2_{V(\mu_n^*)^+} \leq \frac{d}{n}.$$

Confidence bound

If each action $x \in \mathcal{X}$ is sampled $\mu_n^*(x)$ times with $n = \frac{2d}{\epsilon^2}\log\left(k\delta^{-1}\right)$, then

$$\max_{x\in\mathcal{X}} \left|\left(\widehat{\theta} - \theta^*\right)^\top x\right| \leq \epsilon, \quad \text{with probability at least } 1-\delta.$$

G-optimal design

If we choose each action x exactly $\mu(x)$ times

$$\sum_{s\leq n} x_s x_s^\top = V(\mu) := \sum_{x\in\mathcal{X}} \mu(x) x x^\top$$

G-optimal design

$$\mu_n^* \in \underset{|\mu|=n}{\operatorname{argmin}} \ \max_{x\in\mathcal{X}} \|x\|^2_{V(\mu)^+} . \qquad \text{(G-optimal design)}$$

fulfills

$$\max_{x\in\mathcal{X}} \ \|x\|^2_{V(\mu_n^*)^+} \leq \frac{d}{n}.$$

Confidence bound

If each action $x \in \mathcal{X}$ is sampled $\mu_n^*(x)$ times with $n = \frac{2d}{\epsilon^2} \log\left(k\delta^{-1}\right)$, then

$$\max_{x\in\mathcal{X}} \left|\left(\widehat{\theta} - \theta^*\right)^\top x\right| \leq \epsilon, \quad \text{with probability at least } 1-\delta.$$

Phased Elimination algorithm

Phased Elimination (Lattimore and Szepesvári, 2020)

Input $\mathcal{X}_1 = \mathcal{X}$.
For $l = 1, 2, \ldots$

- $\epsilon_l \leftarrow 2^{-l}$, $\quad n_l \leftarrow \dfrac{2d}{\epsilon_l^2} \log\left(kl(l+1)T\right)$
- $\mathcal{X}_{l+1} \leftarrow$G-Explore-And-Eliminate$(\mathcal{X}_l, n_l, \epsilon_l)$

End For

G-Explore-And-Eliminate$(\mathcal{X}_l, n_l, \epsilon_l)$

- sample $\mu^*_{n_l}(x)$ times each action $x \in \mathcal{X}_l$
- compute OLS estimator $\widehat{\theta}$

Return $\mathcal{X}_l \setminus \left\{x \in \mathcal{X}_l : x^\top \widehat{\theta} + \epsilon_l \leq \max_{x \in \mathcal{X}_l} x^\top \widehat{\theta} - \epsilon_l\right\}$

Phased Elimination algorithm

PHASED ELIMINATION (Lattimore and Szepesvári, 2020)

Input $\mathcal{X}_1 = \mathcal{X}$.
For $l = 1, 2, \dots$

- $\epsilon_l \leftarrow 2^{-l}, \quad n_l \leftarrow \dfrac{2d}{\epsilon_l^2} \log\left(kl(l+1)T\right)$
- $\mathcal{X}_{l+1} \leftarrow$ G-EXPLORE-AND-ELIMINATE$(\mathcal{X}_l, n_l, \epsilon_l)$

End For

G-EXPLORE-AND-ELIMINATE$(\mathcal{X}_l, n_l, \epsilon_l)$

- sample $\mu^*_{n_l}(x)$ times each action $x \in \mathcal{X}_l$
- compute OLS estimator $\widehat{\theta}$

Return $\mathcal{X}_l \setminus \left\{ x \in \mathcal{X}_l : x^\top \widehat{\theta} + \epsilon_l \leq \max_{x \in \mathcal{X}_l} x^\top \widehat{\theta} - \epsilon_l \right\}$

Phased Elimination algorithm

Gaps

Gaps: $\Delta_x = (x^* - x)^\top \theta^*$

Minimal gap: $\Delta_{\min} = \min \{\Delta_x : \Delta_x > 0\}$

Theorem

Asymptotic regret: $R_T \lesssim d\Delta_{\min}^{-1} \log(T)$

The worst case regret: $R_T^* \lesssim C\sqrt{dT \log(kT)}$.

Biased linear bandits

Biased Linear Bandits

Our setting

At each round $t = 1, \dots, T$

- the agent chooses an action $x_t \in \mathcal{X} \subset \mathbb{R}^d$, whose sensitive attribute is $z_{x_t} \in \{-1, 1\}$
- she receives the unobserved reward $x_t^\top \gamma^*$;
- she observes the biased feedback

$$
\begin{aligned}
y_t &= x_t^\top \gamma^* + z_{x_t}\omega^* + \xi_t \\
&= a_{x_t}^\top \theta^* + \xi_t, \qquad \text{where } a_x = \begin{pmatrix} x \\ z_x \end{pmatrix}, \ \theta^* = \begin{pmatrix} \gamma^* \\ \omega^* \end{pmatrix}
\end{aligned}
$$

Objective

Minimize the regret $R_T = \mathbb{E}\left[\sum_{t \le T} (x^* - x_t)^\top \gamma^*\right]$, where $x^* \in \operatorname{argmax}_{x \in \mathcal{X}} x^\top \gamma^*$.

Biased Linear Bandits

Our setting

At each round $t = 1, \dots, T$

- the agent chooses an action $x_t \in \mathcal{X} \subset \mathbb{R}^d$, whose sensitive attribute is $z_{x_t} \in \{-1, 1\}$
- she receives the unobserved reward $x_t^\top \gamma^*$;
- she observes the biased feedback

$$
\begin{aligned}
y_t &= x_t^\top \gamma^* + z_{x_t} \omega^* + \xi_t \\
&= a_{x_t}^\top \theta^* + \xi_t, \qquad \text{where } a_x = \begin{pmatrix} x \\ z_x \end{pmatrix}, \ \theta^* = \begin{pmatrix} \gamma^* \\ \omega^* \end{pmatrix}
\end{aligned}
$$

Objective

Minimize the regret $R_T = \mathbb{E}\Big[\sum_{t \le T} (x^* - x_t)^\top \gamma^*\Big]$, where $x^* \in \operatorname{argmax}_{x \in \mathcal{X}} x^\top \gamma^*$.

A first naive idea

Bias issue

We have the linear model $y = a_x^\top \theta^* + \xi$, but the reward is

$$x^\top \gamma^* = a_x^\top \theta^* - z_x \omega^*.$$

Naive Debiased G-Explore-And-Eliminate($\mathcal{X}_l, n_l, \epsilon_l$)

- $\mu^*_{n_l} \leftarrow$ G-optimal design($\{a_x : x \in \mathcal{X}_l\}, n_l$)
- sample $\mu^*_{n_l}(x)$ times each action $x \in \mathcal{X}_l$
- compute OLS estimator $\widehat{\theta} = \begin{pmatrix} \widehat{\gamma} \\ \widehat{\omega} \end{pmatrix}$

Return $\mathcal{X}_l \setminus \left\{ x \in \mathcal{X}_l : x^\top \widehat{\gamma} + \epsilon_l \leq \max_{x \in \mathcal{X}_l} x^\top \widehat{\gamma} - \epsilon_l \right\}$

A first naive idea

Bias issue

We have the linear model $y = a_x^\top \theta^* + \xi$, but the reward is

$$x^\top \gamma^* = a_x^\top \theta^* - z_x \omega^*.$$

Naive $\textsc{Debiased G-Explore-And-Eliminate}(\mathcal{X}_l, n_l, \epsilon_l)$

- $\mu^*_{n_l} \leftarrow$ G-optimal design$(\{a_x : x \in \mathcal{X}_l\}, n_l)$
- sample $\mu^*_{n_l}(x)$ times each action $x \in \mathcal{X}_l$
- compute OLS estimator $\widehat{\theta} = \begin{pmatrix} \widehat{\gamma} \\ \widehat{\omega} \end{pmatrix}$

Return $\mathcal{X}_l \setminus \left\{ x \in \mathcal{X}_l : x^\top \widehat{\gamma} + \epsilon_l \leq \max_{x \in \mathcal{X}_l} x^\top \widehat{\gamma} - \epsilon_l \right\}$

Failure of the naive idea

Caveat

We have with probability at least $1-\delta$

$$\max_{x\in\mathcal{X}_l}\left|\left(\widehat{\theta}-\theta^*\right)^\top a_x\right|\leq\epsilon_l.$$

But, we have no control on the reward ☹

$$\max_{x\in\mathcal{X}_l}\left|\left(\widehat{\gamma}-\gamma^*\right)^\top x\right|\leq ??$$

Remedy

We need an additional step of optimal-designed estimation of the bias ω^*.

Failure of the naive idea

Caveat

We have with probability at least $1 - \delta$

$$\max_{x \in \mathcal{X}_l} \left| \left(\widehat{\theta} - \theta^* \right)^\top a_x \right| \le \epsilon_l .$$

But, we have no control on the reward ☹

$$\max_{x \in \mathcal{X}_l} \left| (\widehat{\gamma} - \gamma^*)^\top x \right| \le ??$$

Remedy

We need an additional step of optimal-designed estimation of the bias ω^*.

Bias estimation

OLS estimation (best unbiased linear estimator)

- Sample $\mu(x)$ times each $x \in \mathcal{X}$
- Compute the OLS estimator $\widehat{\theta} = \begin{pmatrix} \widehat{\gamma} \\ \widehat{\omega} \end{pmatrix}$

Then,

$$\mathbb{P}\left(|\widehat{\omega} - \omega^*| \leq \sqrt{2 \, \|e_{d+1}\|^2_{V(\mu)^+} \log(1/\delta)} \right) \geq 1 - \delta,$$

where $V(\mu) = \sum_x \mu(x) a_x a_x^\top$ and $e_{d+1} = (0, \ldots, 0, 1)$.

Regret for bias estimation

When we sample $\mu(x)$ times each $x \in \mathcal{X}$, we suffer the regret

$$\sum_{x \in \mathcal{X}} \mu(x) \Delta_x, \quad \text{where} \quad \Delta_x = (x^* - x)^\top \gamma^*.$$

Bias estimation

OLS estimation (best unbiased linear estimator)

- Sample $\mu(x)$ times each $x \in \mathcal{X}$
- Compute the OLS estimator $\widehat{\theta} = \begin{pmatrix} \widehat{\gamma} \\ \widehat{\omega} \end{pmatrix}$

Then,

$$\mathbb{P}\left(|\widehat{\omega} - \omega^*| \leq \sqrt{2 \left\| e_{d+1} \right\|^2_{V(\mu)^+} \log\left(1/\delta\right)}\right) \geq 1 - \delta,$$

where $V(\mu) = \sum_x \mu(x) a_x a_x^\top$ and $e_{d+1} = (0, \ldots, 0, 1)$.

Regret for bias estimation

When we sample $\mu(x)$ times each $x \in \mathcal{X}$, we suffer the regret

$$\sum_{x \in \mathcal{X}} \mu(x) \Delta_x, \quad \text{where} \quad \Delta_x = (x^* - x)^\top \gamma^*.$$

Δ-optimal design

Δ-optimal design

For $\Delta = (\Delta_x)_{x\in\mathcal{X}}$, we introduce

$$\mu^{\Delta} = \operatorname*{argmin}_{\substack{\mu \in \mathcal{M}^{\mathcal{X}}_{e_{d+1}} \text{ s.t.} \\ \|e_{d+1}\|^2_{V(\mu)^+} \leq 1}} \sum_{x\in\mathcal{X}} \mu(x)\Delta_x \qquad (\Delta\text{-optimal design})$$

Δ-optimal design

Δ-optimal design

For $\Delta = (\Delta_x)_{x \in \mathcal{X}}$, we introduce

$$\sigma^{-1}\mu^{\Delta} = \underset{\substack{\mu \in \mathcal{M}^{\mathcal{X}}_{e_{d+1}} \text{ s.t.} \\ \|e_{d+1}\|^2_{V(\mu)^+} \leq \sigma^2}}{\text{argmin}} \sum_{x \in \mathcal{X}} \mu(x)\Delta_x \qquad (\Delta\text{-optimal design})$$

Regret for Δ-optimal bias estimation

Regret for bias estimation

The minimal regret for estimating the bias ω^* with precision ϵ and confidence $1-\delta$ is

$$\text{minimal regret for bias estimation} = \frac{2\kappa(\Delta)\log(\delta^{-1})}{\epsilon^2}\ ,$$

where

$$\kappa(\Delta) = \sum_{x\in\mathcal{X}} \mu^{\Delta}(x)\Delta_x$$

characterizes the difficulty of bias estimation in our setting.

In practice

In practice, Δ_x is unknown, we have to rely on some (upper) estimates $\widehat{\Delta}_x$.

Regret for Δ-optimal bias estimation

Regret for bias estimation

The minimal regret for estimating the bias ω^* with precision ϵ and confidence $1-\delta$ is

$$\text{minimal regret for bias estimation} = \frac{2\kappa(\Delta)\log(\delta^{-1})}{\epsilon^2}\ ,$$

where

$$\kappa(\Delta) = \sum_{x\in\mathcal{X}} \mu^{\Delta}(x)\Delta_x$$

characterizes the difficulty of bias estimation in our setting.

In practice

In practice, Δ_x is unknown, we have to rely on some (upper) estimates $\widehat{\Delta}_x$.

Fair Phased Elimination

3 main ingredients

1. Actions can be compared within a group:
   we apply G-Explore-And-Eliminate within a group
2. bias correction based on $\widehat{\Delta}$-optimal design
3. bias estimation breaking criterion: to avoid a too high regret

Fair Phased Elimination algorithm

Fair Phased Elimination

Input $\mathcal{Z} = \{-1, +1\}$, and $\mathcal{X}_1^z = \{x \in \mathcal{X} : z_x = z\}$ for $z \in \mathcal{Z}$

For $l = 1, 2, \ldots$

- $\epsilon_l \leftarrow 2^{-l}$, $n_l \leftarrow \frac{2(d+1)}{\epsilon_l^2} \log\left(\frac{kl(l+1)}{\delta}\right)$, $m_l \leftarrow \frac{2}{\epsilon_l^2} \log\left(\frac{l(l+1)}{\delta}\right)$
- **For** $z \in \mathcal{Z}$
  - $\mathcal{X}_{l+1}^z, \widehat{\theta}_l^{(z)} \leftarrow$G-Explore-And-Eliminate$(\mathcal{X}_l^z, n_l, \epsilon_l)$
- **If** $\mathcal{Z} = \{-1, +1\}$
  - **If** $\epsilon_l \leq \left(\kappa(\widehat{\Delta}^l) \log(T)/T\right)^{1/3}$, **then** break and sample best empirical action for remaining time
  - $\widehat{\omega}_l \leftarrow \Delta$-Explore$(\widehat{\Delta}^l, m_l)$
  - $\widehat{m}_x \leftarrow a_x^\top \widehat{\theta}_l^{(z)} - z\widehat{\omega}_l$ for $x \in \mathcal{X}_l^{(-1)} \cup \mathcal{X}_l^{(1)}$, update $\widehat{\Delta}^l$
  - **If** $\exists z \in \mathcal{Z}$ s.t. $\max_{x \in \mathcal{X}_l^{(z)}} \widehat{m}_x \geq \max_{x \in \mathcal{X}_l^{(-z)}} \widehat{m}_x + 4\epsilon_l$ **then** $\mathcal{Z} \leftarrow \{z\}$

Optimal Regret for Biased Linear Bandits

Geometry of worst-case regret

Theorem (Gaucher, Carpentier, Giraud, 2022)

Define

$$\kappa_* = \min_{\substack{\mu \in \mathcal{M}^{\mathcal{X}}_{e_{d+1}} \text{ s.t.} \\ \|e_{d+1}\|^2_{V(\mu)^+} \leq 1}} \max_{\theta : |a_x^\top \theta| \leq 1} \sum_{x \in \mathcal{X}} \mu(x) \Delta_x(\theta).$$

Then, Fair Phased Elimination *algorithm fulfills*

$$R_T \leq C \, \kappa_*^{1/3} T^{2/3} \log(T)^{1/3}, \qquad \textit{for} \quad T \geq T_{k,d,\kappa_*}.$$

Remarks

- Matching lower bound up to a $\log(T)^{1/3}$;
- $\kappa_*^{1/3}$ captures the dependency on the geometry of the set of actions;
- Regret in $\tilde{\Theta}(T^{2/3})$ instead of $\tilde{\Theta}(T^{1/2})$ is the price for debiasing the rewards.

Geometry of worst-case regret

Theorem (Gaucher, Carpentier, Giraud, 2022)

Define

$$\kappa_* = \min_{\substack{\mu \in \mathcal{M}^{\mathcal{X}}_{e_{d+1}} \text{ s.t.} \\ \|e_{d+1}\|^2_{V(\mu)^+} \leq 1}} \max_{\theta : |a_x^\top \theta| \leq 1} \sum_{x \in \mathcal{X}} \mu(x) \Delta_x(\theta).$$

Then, Fair Phased Elimination *algorithm fulfills*

$$R_T \leq C\, \kappa_*^{1/3} T^{2/3} \log(T)^{1/3}, \qquad \textit{for} \quad T \geq T_{k,d,\kappa_*}.$$

Remarks

- Matching lower bound up to a $\log(T)^{1/3}$;
- $\kappa_*^{1/3}$ captures the dependency on the geometry of the set of actions;
- Regret in $\tilde{\Theta}(T^{2/3})$ instead of $\tilde{\Theta}(T^{1/2})$ is the price for debiasing the rewards.

Geometry of worst-case regret

Theorem (Gaucher, Carpentier, Giraud, 2022)

Define

$$\kappa_* = \min_{\substack{\mu \in \mathcal{M}^{\mathcal{X}}_{e_{d+1}} \ s.t. \\ \|e_{d+1}\|^2_{V(\mu)^+} \leq 1}} \ \max_{\theta : |a_x^\top \theta| \leq 1} \sum_{x \in \mathcal{X}} \mu(x) \Delta_x(\theta).$$

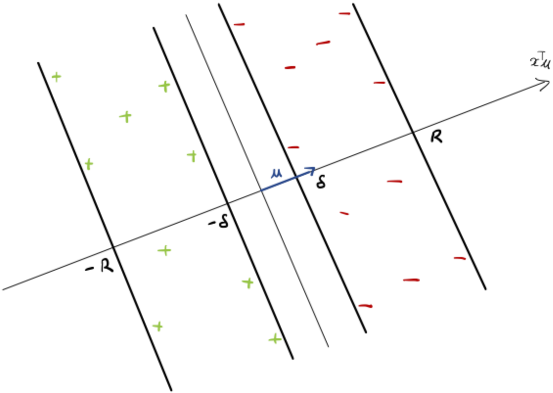
Then, Fair Phased Elimination *algorithm fulfills*

$$R_T \leq C \, \kappa_*^{1/3} T^{2/3} \log(T)^{1/3}, \qquad \textit{for} \quad T \geq T_{k,d,\kappa_*}.$$

Remarks

- Matching lower bound up to a $\log(T)^{1/3}$;
- $\kappa_*^{1/3}$ captures the dependency on the geometry of the set of actions;
- Regret in $\tilde{\Theta}(T^{2/3})$ instead of $\tilde{\Theta}(T^{1/2})$ is the price for debiasing the rewards.

Geometry of bias estimation



Lemma

$$\kappa^* = \Delta_{\max} \left(\frac{R+\delta}{R-\delta} \right)^2$$

with the largest $\delta/R \in [0,1]$ such that a δ-separation as above exists

Δ-dependent regret bound

$$\Delta_{\min} := \min_{x \neq x^*} (x^* - x)^\top \gamma^* \qquad \text{and} \qquad \Delta_{\neq} := \min_{z_x \neq z_{x^*}} (x^* - x)^\top \gamma^*.$$

Theorem (Gaucher, Carpentier, Giraud, 2022)

Fair Phased Elimination *algorithm fulfills*

$$R_T \leq C \left(\frac{d}{\Delta_{\min}} + \frac{\kappa(\Delta \vee \Delta_{\neq} \vee \varepsilon_T)}{\Delta_{\neq}^2} \right) \log(T), \qquad \textit{for} \quad T \geq k \vee e^{d\Delta_{\min}}$$

where $\varepsilon_T = (\kappa_* \log(T)/T)^{1/3}$.

Comments

- Some matching lower bounds;
- $\frac{d \log(T)}{\Delta_{\min}}$ is the (worst gap-dependent) regret of the classical linear bandit;
- $\frac{\kappa(\Delta) \log(T)}{\Delta_{\neq}^2}$ is the price for debiasing the rewards.

Δ-dependent regret bound

$$\Delta_{\min} := \min_{x \neq x^*} (x^* - x)^\top \gamma^* \qquad \text{and} \qquad \Delta_{\neq} := \min_{z_x \neq z_{x^*}} (x^* - x)^\top \gamma^*.$$

Theorem (Gaucher, Carpentier, Giraud, 2022)

Fair Phased Elimination *algorithm fulfills*

$$R_T \leq C \left(\frac{d}{\Delta_{\min}} + \frac{\kappa\big(\Delta \vee \Delta_{\neq} \vee \varepsilon_T\big)}{\Delta_{\neq}^2} \right) \log(T), \qquad \textit{for} \quad T \geq k \vee e^{d\Delta_{\min}}$$

where $\varepsilon_T = (\kappa_* \log(T)/T)^{1/3}$.

Comments

- Some matching lower bounds;
- $\frac{d \log(T)}{\Delta_{\min}}$ is the (worst gap-dependent) regret of the classical linear bandit;
- $\frac{\kappa(\Delta) \log(T)}{\Delta_{\neq}^2}$ is the price for debiasing the rewards.

Δ-dependent regret bound

$$\Delta_{\min} := \min_{x \neq x^*} (x^* - x)^\top \gamma^* \qquad \text{and} \qquad \Delta_{\neq} := \min_{z_x \neq z_{x^*}} (x^* - x)^\top \gamma^*.$$

Theorem (Gaucher, Carpentier, Giraud, 2022)

Fair Phased Elimination *algorithm fulfills*

$$R_T \leq C \left(\frac{d}{\Delta_{\min}} + \frac{\kappa(\Delta \vee \Delta_{\neq} \vee \varepsilon_T)}{\Delta_{\neq}^2} \right) \log(T), \qquad \textit{for} \quad T \geq k \vee e^{d\Delta_{\min}}$$

where $\varepsilon_T = (\kappa_* \log(T)/T)^{1/3}$.

Comments

- Some matching lower bounds;
- $\frac{d \log(T)}{\Delta_{\min}}$ is the (worst gap-dependent) regret of the classical linear bandit;
- $\frac{\kappa(\Delta) \log(T)}{\Delta_{\neq}^2}$ is the price for debiasing the rewards.

Take home message

In biased linear bandit problems

- In the worst case, the regret can be $\tilde{\Theta}(T^{2/3})$ instead of $\tilde{\Theta}(\sqrt{T})$.
  The geometric dependence is captured by the largest δ-separation.
- In gap-depend worst case:
  - an additional $\frac{\kappa(\Delta)\log(T)}{\Delta_{\neq}^2}$ term shows up
  - can be as easy as classical bandit if $\frac{\kappa(\Delta)\log(T)}{\Delta_{\neq}^2} \leq \frac{d\log(T)}{\Delta_{\min}}$