CONTRIBUTIONS TO STOCHASTIC BANDITS AND LINK PREDICTION PROBLEMS

Solenne Gaucher

27 juin 2022

Laboratoire de Mathématiques d'Orsay, Université Paris-Saclay

Outline

1. Link prediction and network estimation
   - Maximum likelihood estimation in the SBM and its variational approximation
   - Robust link prediction and outlier detection
2. Stochastic bandits
   - Finite continuum-armed bandits
   - Biased linear bandits

Outline

1. Link prediction and network estimation
   - **Maximum likelihood estimation in the SBM** and its variational approximation
   - Robust link prediction and outlier detection
2. Stochastic bandits
   - **Finite continuum-armed bandits**
   - **Biased linear bandits**

MAXIMUM LIKELIHOOD ESTIMATION FOR LINK PREDICTION

Motivation

Definition : A network is given by

- a set of nodes;
- a set of edges linking these nodes.

Networks are used to model complex systems of interactions :

- social networks;
- protein-protein interactions;
- ecological networks;
- ...

Definitions

Consider **undirected**, **unweighted** network with **no self-loop**.

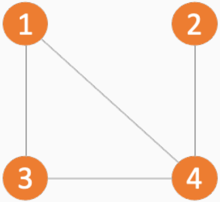
The network is described by

- a set of n nodes $\{1, ..., n\}$;
- a set of edges $\mathcal{E} = \{\{i, j\}, 1 \leq i, j \leq n\}$.

The **adjacency matrix** is given by

$$\mathbf{A}_{ij} = \begin{cases} 1 & \text{if } \{i, j\} \in \mathcal{E} \\ 0 & \text{otherwise.} \end{cases}$$

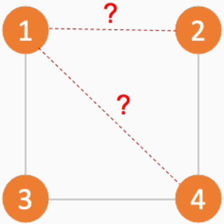
Example



$$\mathbf{A} = \left\{ \begin{array}{cccc} 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{array} \right\}$$

Missing observations

Missing observations on the presence or absence of edges.



$$\mathbf{A} = \begin{Bmatrix} 0 & ? & 1 & ? \\ ? & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ ? & 1 & 1 & 0 \end{Bmatrix} \quad \Omega = \begin{Bmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 \end{Bmatrix}$$

Model

The adjacency matrix $\mathbf{A}$ is random :

$$\mathbf{A}_{ij}|\Theta^*_{ij} \overset{i.i.d}{\sim} \text{Bernoulli}(\Theta^*_{ij}).$$

$\Theta^* \in [0,1]^{n\times n}$ is the matrix of connection probabilities.

The **sampling matrix** is random :

$$\Omega_{ij} \overset{i.i.d}{\sim} \text{Bernoulli}(\Pi_{ij}).$$

Objective : Estimate Θ^* for

- network denoising;
- link prediction.

Need assumptions on the structure of Θ^* !

Model

The adjacency matrix $\mathbf{A}$ is random :

$$\mathbf{A}_{ij}|\Theta^*_{ij} \overset{i.i.d}{\sim} \text{Bernoulli}(\Theta^*_{ij}).$$

$\Theta^* \in [0,1]^{n\times n}$ is the matrix of connection probabilities.

The **sampling matrix** is random :

$$\Omega_{ij} \overset{i.i.d}{\sim} \text{Bernoulli}(\Pi_{ij}).$$

Objective : Estimate Θ^* for

- network denoising;
- link prediction.

Need assumptions on the structure of Θ^*!

Model

The adjacency matrix $\mathbf{A}$ is random :

$$\mathbf{A}_{ij}|\Theta^*_{ij} \overset{i.i.d}{\sim} \text{Bernoulli}(\Theta^*_{ij}).$$

$\Theta^* \in [0,1]^{n\times n}$ is the matrix of connection probabilities.

The **sampling matrix** is random :

$$\Omega_{ij} \overset{i.i.d}{\sim} \text{Bernoulli}(\Pi_{ij}).$$

Objective : Estimate Θ^* for

- network denoising;
- link prediction.

Need assumptions on the structure of Θ^*!

Community structures in networks

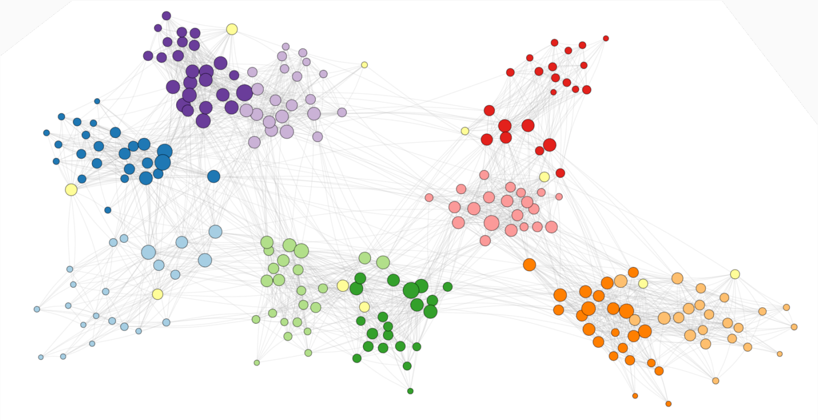


Figure : Network of interactions within a primary school, Stehlé et al. (2011).

The stochastic block model

Stochastic block model with k communities :

- each node i belongs to a community $z_i^* \in \{1, ..., k\}$;
- $z^* \in \{1, ..., k\}^n$ is the vector of communities ;
- $\mathbf{Q}^* \in [0, 1]^{k \times k}$ is the matrix of connection probabilities of the communities ;
- for all pairs of nodes (i, j), $i \neq j$:

$$\mathbb{P}\left(\mathbf{A}_{ij} = 1 | z_i^* = a, z_j^* = b\right) = \mathbf{Q}^*_{a,b}.$$

The stochastic block model

Stochastic block model with k communities :

- each node i belongs to a community $z_i^* \in \{1, ..., k\}$;
- $z^* \in \{1, ..., k\}^n$ is the vector of communities ;
- $\mathbf{Q}^* \in [0, 1]^{k \times k}$ is the matrix of connection probabilities of the communities ;
- for all pairs of nodes (i, j), $i \neq j$:

$$\Theta(\mathbf{Q}^*, z^*)_{ij} = \mathbb{P}\left(\mathbf{A}_{ij} = 1 | z_i^* = a, z_j^* = b\right) = \mathbf{Q}^*_{a,b}.$$

The stochastic block model

Stochastic block model with k communities :

- each node i belongs to a community $z_i^* \in \{1, ..., k\}$;
- $z^* \in \{1, ..., k\}^n$ is the vector of communities ;
- $\mathbf{Q}^* \in [0, 1]^{k \times k}$ is the matrix of connection probabilities of the communities ;
- for all pairs of nodes (i, j), $i \neq j$:

$$\Theta(\mathbf{Q}^*, z^*)_{ij} = \mathbb{P}\left(\mathbf{A}_{ij} = 1 | z_i^* = a, z_j^* = b\right) = \mathbf{Q}^*_{a,b}.$$

Sparse networks : $\|\mathbf{Q}^*\|_\infty \leq \rho_n$.

Gao et al. (2015), Klopp et al. (2017), Gao et al. (2016)

Least square estimation for uniform sampling ($\Pi_{ij} = p$) :

$$\left(\widehat{\mathbf{Q}}^{LS}, \hat{z}^{LS}\right) \in \underset{\mathbf{Q}\in[0,\rho_n]^{k\times k}_{sym}, z\in\{1,..,k\}^n}{\operatorname{argmin}} \|\Theta(\mathbf{Q}, z)\|_F^2 - \frac{2}{p}\sum_{i<j}\Omega_{ij}\mathbf{A}_{ij}\mathbf{Q}_{z_i,z_j}.$$

Convergence rate : With large probability,

$$\left\|\Theta^* - \Theta(\widehat{\mathbf{Q}}^{LS}, \hat{z}^{LS})\right\|_F^2 \leq C\frac{\rho_n}{p}\left(k^2 + n\log(k)\right).$$

Minimax optimal if $p\rho_n \geq \frac{\log(k)}{n}$.

Least square estimator cannot be computed in polynomial time!

Previous results

Gao et al. (2015), Klopp et al. (2017), Gao et al. (2016)

Least square estimation for uniform sampling ($\Pi_{ij} = p$) :

$$\left(\widehat{\mathbf{Q}}^{LS}, \hat{z}^{LS}\right) \in \underset{\mathbf{Q}\in[0,\rho_n]^{k\times k}_{sym}, z\in\{1,..,k\}^n}{\operatorname{argmin}} \|\Theta(\mathbf{Q}, z)\|_F^2 - \frac{2}{p}\sum_{i<j}\Omega_{ij}\mathbf{A}_{ij}\mathbf{Q}_{z_i,z_j}.$$

Convergence rate : With large probability,

$$\left\|\Theta^* - \Theta(\widehat{\mathbf{Q}}^{LS}, \hat{z}^{LS})\right\|_F^2 \leq C\frac{\rho_n}{p}\left(k^2 + n\log(k)\right).$$

Minimax optimal if $p\rho_n \geq \frac{\log(k)}{n}$.

Least square estimator cannot be computed in polynomial time!

Previous results

Gao et al. (2015), Klopp et al. (2017), Gao et al. (2016)

Least square estimation for uniform sampling ($\Pi_{ij} = p$) :

$$\left(\widehat{\mathbf{Q}}^{LS}, \hat{z}^{LS}\right) \in \underset{\mathbf{Q}\in[0,\rho_n]^{k\times k}_{sym}, z\in\{1,..,k\}^n}{\operatorname{argmin}} \|\Theta(\mathbf{Q}, z)\|_F^2 - \frac{2}{p}\sum_{i<j}\Omega_{ij}\mathbf{A}_{ij}\mathbf{Q}_{z_i,z_j}.$$

Convergence rate : With large probability,

$$\left\|\Theta^* - \Theta(\widehat{\mathbf{Q}}^{LS}, \hat{z}^{LS})\right\|_F^2 \le C\frac{\rho_n}{p}\left(k^2 + n\log(k)\right).$$

Minimax optimal if $p\rho_n \ge \frac{\log(k)}{n}$.

Least square estimator cannot be computed in polynomial time!

Maximum likelihood estimation

Assumptions : For all $i \neq j$, $0 < \gamma_n \leq \Theta^*_{i,j} \leq \rho_n < 1$.

Maximum likelihood estimator :

$$\left(\widehat{\mathbf{Q}}^{ML}, \hat{z}^{ML}\right) \in \underset{\mathbf{Q} \in [\gamma_n, \rho_n]^{k\times k}_{sym}, z \in \{1,..,k\}^n}{\operatorname{argmax}} \mathcal{L}_{\Omega}(\mathbf{Q}, z)$$

$$\mathcal{L}_{\Omega}(\mathbf{Q}, z) = \sum_{i<j} \Omega_{ij} \left(\mathbf{A}_{ij} \log(\mathbf{Q}_{z_i, z_j}) + (1 - \mathbf{A}_{ij}) \log(1 - \mathbf{Q}_{z_i, z_j})\right).$$

Maximum likelihood estimator cannot be computed in polynomial time... **Variational approximation** is used in practice.

Maximum likelihood estimation

Assumptions : For all $i \neq j$, $0 < \gamma_n \leq \Theta^*_{i,j} \leq \rho_n < 1$.

Maximum likelihood estimator :

$$\left(\widehat{\mathbf{Q}}^{ML}, \hat{z}^{ML}\right) \in \underset{\mathbf{Q} \in [\gamma_n, \rho_n]^{k \times k}_{sym}, z \in \{1,..,k\}^n}{\operatorname{argmax}} \mathcal{L}_{\Omega}(\mathbf{Q}, z)$$

$$\mathcal{L}_{\Omega}(\mathbf{Q}, z) = \sum_{i<j} \Omega_{ij} \left(\mathbf{A}_{ij} \log(\mathbf{Q}_{z_i, z_j}) + (1 - \mathbf{A}_{ij}) \log(1 - \mathbf{Q}_{z_i, z_j}) \right).$$

Maximum likelihood estimator cannot be computed in polynomial time... **Variational approximation** is used in practice.

Maximum likelihood estimation

Assumptions : For all $i \neq j$, $0 < \gamma_n \leq \Theta^*_{i,j} \leq \rho_n < 1$.

Maximum likelihood estimator :

$$\left(\widehat{\mathbf{Q}}^{ML}, \hat{z}^{ML}\right) \in \underset{\mathbf{Q}\in[\gamma_n,\rho_n]^{k\times k}_{sym}, z\in\{1,..,k\}^n}{\operatorname{argmin}} \mathcal{K}_\Omega\left(\mathbf{A}, \Theta(\mathbf{Q}, z)\right)$$

$$\mathcal{K}_\Omega\left(\mathbf{A}, \Theta\right) = \sum_{i<j} \Omega_{ij} kl\left(\mathbf{A}_{ij}, \Theta_{ij}\right).$$

Maximum likelihood estimator cannot be computed in polynomial time... **Variational approximation** is used in practice.

Results

Theorem (G., Klopp (2021))

If $\rho_n \gg n^{-1}$*, with probability at least* $1 - 9\exp\left(-C\rho_n n\log(k)\right)$*,*

$$\|\Theta^* - \widehat{\Theta}\|_{\Pi}^2 \leq C'\rho_n\left(k^2 + n\log(k)\right) \times \left(\frac{\rho_n}{(1-\rho_n)\wedge\gamma_n}\right)^2.$$

where $\|\Theta^* - \widehat{\Theta}\|_{\Pi}^2 = \sum_{i<j} \Pi_{ij}\left(\Theta^*_{ij} - \widehat{\Theta}_{ij}\right)^2$

Assume uniform sampling ($\Pi_{ij} = p$), and that $\mathbf{Q}^* = \rho_n \mathbf{Q}^0$ with $0 < \mathbf{Q}^0_{ab} < 1$.

Corollary (G., Klopp (2021))

If $\rho_n \gg n^{-1}$, with probability at least $1 - 9\exp\left(-C\rho_n n \log(k)\right)$,

$$\|\Theta^* - \widehat{\Theta}\|_F^2 \leq C_{\mathbf{Q}^0} \frac{\rho_n}{p} \left(k^2 + n\log(k)\right).$$

Then, maximum likelihood estimation is minimax optimal.

SBM approximation of regular graphon

Sparse graphon model :
$\Theta^*_{i,j} = \rho_n W(\zeta_i, \zeta_j)$, where $\zeta_i \overset{i.i.d.}{\sim} \mathcal{U}([0,1])$.

Approximate W using a SBM with k communities.

Theorem (G., Klopp (2021))

If $0 < c < W(x,y) < 1$ and $\rho_n \gg n^{-1}$, with probability at least $1 - 9\exp\left(-C\rho_n n \log(k)\right)$,

$$\|\Theta^* - \widehat{\Theta}\|^2_{\Pi} \leq C_c \rho_n \left((k^2 + n\log(k)) + \mathcal{K}_{\Pi}\left(\Theta^*, \Theta(\widetilde{\mathbf{Q}}, \tilde{z})\right)\right).$$

$$\left(\widetilde{\mathbf{Q}}, \tilde{z}\right) \in \underset{\mathbf{Q}\in[\gamma_n,\rho_n]^{k\times k}_{sym}, z\in\{1,..,k\}^n}{\operatorname{argmin}} \mathcal{K}_{\Pi}(\Theta^*, \Theta(\mathbf{Q}, z)).$$

SBM approximation of regular graphon

Sparse graphon model :
$\Theta^*_{i,j} = \rho_n W(\zeta_i, \zeta_j)$, where $\zeta_i \overset{i.i.d.}{\sim} \mathcal{U}([0,1])$.

Approximate W using a SBM with k communities.

Theorem (G., Klopp (2021))

If $0 < c < W(x,y) < 1$ and $\rho_n \gg n^{-1}$, with probability at least $1 - 9\exp\left(-C\rho_n n \log(k)\right)$,

$$\|\Theta^* - \widehat{\Theta}\|_{\Pi}^2 \leq C_c \rho_n \left(\left(k^2 + n\log(k)\right) + \mathcal{K}_{\Pi}\left(\Theta^*, \Theta(\widetilde{\mathbf{Q}}, \tilde{z})\right)\right).$$

$$\left(\widetilde{\mathbf{Q}}, \tilde{z}\right) \in \underset{\mathbf{Q}\in[\gamma_n,\rho_n]^{k\times k}_{sym}, z\in\{1,..,k\}^n}{\operatorname{argmin}} \mathcal{K}_{\Pi}\left(\Theta^*, \Theta(\mathbf{Q}, z)\right).$$

Conclusion

- The maximum likelihood estimator cannot be computed in polynomial time, but its computationally efficient variational approximations are used.
- When Θ^* has entries of the same order of magnitude, MLE is minimax optimal.
- MLE is adaptive to the sampling scheme Π.
- When the network follows a smooth graphon model, we can use the stochastic block model as an approximation.

FINITE CONTINUUM-ARMED BANDITS

Bandits for sequential decision-making problems

A sequential decision problem : At each round $t = 1, \dots, T$

- the agent chooses an action $k_t \in \{1, ..., N\}$ based on the observations collected so far;
- she receives a reward y_t such that $\mathbb{E}\left[y_t | k_t\right] = m_{k_t}$.

Goal : Maximize the cumulative reward :

$$\mathbb{E}\Big[\sum_{t \le T} m_{k_t}\Big].$$

Bandits for sequential decision-making problems

A sequential decision problem : At each round $t = 1, \dots, T$

- the agent chooses an action $k_t \in \{1, \dots, N\}$ based on the observations collected so far ;
- she receives a reward y_t such that $\mathbb{E}[y_t | k_t] = m_{k_t}$.

Goal : Minimize the regret :

$$R_T = T m_{k^*} - \mathbb{E}\Big[\sum_{t \le T} m_{k_t}\Big], \quad \text{where} \quad k^* \in \operatorname*{argmax}_{k \le N} m_k.$$

Exploration-exploitation trade-off

Two bandit problems

Finite continuum-armed bandits :

- the agent has access to a set of actions with covariates ;
- each action can only be chosen **once**.

$\rightarrow$ Motivation : allocation of a finite budget between competing candidates.

Biased linear bandits :

- the agent has access to a set of actions with covariates ;
- the feedback for choosing an action is **biased** against a group of actions.

$\rightarrow$ Motivation : concerns regarding unfair evaluations.

Finite continuum-armed bandit

An agent is presented with a set of actions $\{a_1, ..., a_N\}$ (we consider $a_i \in [0, 1]$).

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

- the agent chooses an action $\phi(t) \in \{1, ..., N\}$ with covariate $a_{\phi(t)}$, that she has not yet chosen;
- she receives the reward $y_t \in [0, 1]$ such that $\mathbb{E}[y_t | a_{\phi(t)} = a] = m(a)$.

Goal : Maximize the cumulative reward.

Variant of the continuum-armed bandit (Kleinberg (2004), Auer et al. (2007), Bubeck et al. (2007)).

Finite continuum-armed bandit

An agent is presented with a set of actions $\{a_1, ..., a_N\}$ (we consider $a_i \in [0, 1]$).

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

- the agent chooses an action $\phi(t) \in \{1, ..., N\}$ with covariate $a_{\phi(t)}$, that she has not yet chosen;
- she receives the reward $y_t \in [0, 1]$ such that $\mathbb{E}[y_t | a_{\phi(t)} = a] = m(a)$.

Goal : Maximize the cumulative reward.

Variant of the continuum-armed bandit (Kleinberg (2004), Auer et al. (2007), Bubeck et al. (2007)).

Preliminary remarks

No-repetition constraint :

- leads to lower cumulative rewards;
- changes the exploration-exploitation trade-off;
- changes the regret.

$$R_T = \sum_{t \leq T} m(a_{\phi^*(t)}) - \sum_{t \leq T} m(a_{\phi(t)}).$$

The oracle strategy ϕ^* is such that

$$m(a_{\phi^*(1)}) \geq m(a_{\phi^*(2)}) \geq ... \geq m(a_{\phi^*(N)}).$$

Preliminary remarks

No-repetition constraint :

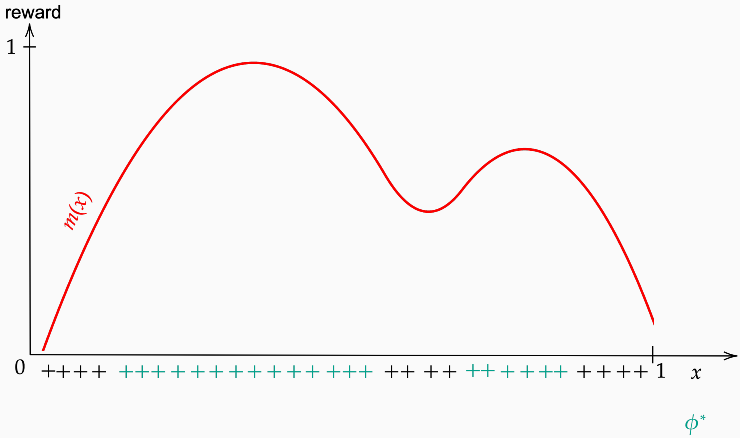
- leads to lower cumulative rewards;
- changes the exploration-exploitation trade-off;
- changes the regret :

$$R_T = \sum_{t \leq T} m(a_{\phi^*(t)}) - \sum_{t \leq T} m(a_{\phi(t)}).$$

The oracle strategy ϕ^* is such that

$$m(a_{\phi^*(1)}) \geq m(a_{\phi^*(2)}) \geq ... \geq m(a_{\phi^*(N)}).$$

Example



Goal : Minimize the regret

$$R_T = \sum_{t \leq T} m(a_{\phi^*(t)}) - \sum_{t \leq T} m(a_{\phi(t)}).$$

The oracle strategy ϕ^* is such that

$$m(a_{\phi^*(1)}) \geq m(a_{\phi^*(2)}) \geq ... \geq m(a_{\phi^*(N)}).$$

Remark :

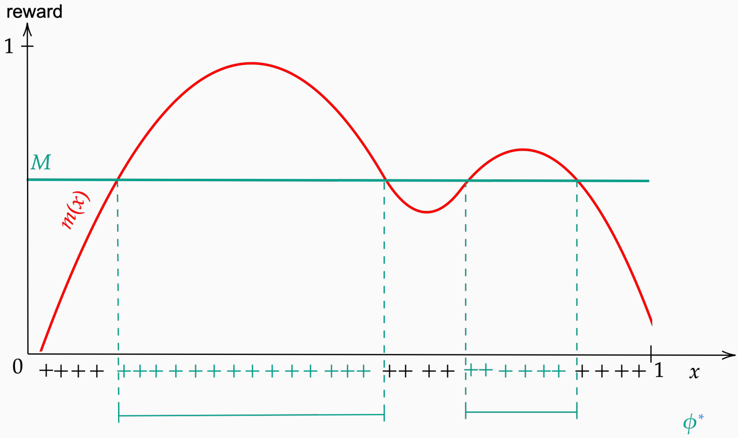
- ϕ^* has decreasing rewards ;
- difficulty of the problem governed by $p = T/N$.

Assumptions

Assumption (A1) : $a_i \overset{i.i.d}{\sim} \mathcal{U}([0,1])$.

The oracle strategy ϕ^* selects actions a such that $m(a) \geq m(a_{\phi^*(T)})$. Under Assumption (A1), $m(a_{\phi^*(T)}) \approx M$, where

$$M = \min \{A : \lambda(\{x : m(x) \geq A\}) < p\}.$$



reward
1
M
m(x)
0
1
x
ϕ*

Assumptions

$$M = \min \{A : \lambda(\{x : m(x) \geq A\}) < p\}.$$

Assumption (A1) : $a_i \overset{i.i.d}{\sim} \mathcal{U}([0,1])$.

Assumptions

$$M = \min \{A : \lambda(\{x : m(x) \geq A\}) < p\}.$$

Assumption (A1) : $a_i \overset{i.i.d}{\sim} \mathcal{U}([0,1])$.

Assumption (A2) : For all $(x, y) \in [0,1]^2$,

$$|m(x) - m(y)| \leq \max\{|M - m(x)|, L\,|x - y|\}.$$

Assumptions

$$M = \min\{A : \lambda(\{x : m(x) \geq A\}) < p\}.$$

Assumption (A1) : $a_i \overset{i.i.d}{\sim} \mathcal{U}([0,1])$.

Assumption (A2) : For all $(x, y) \in [0,1]^2$,

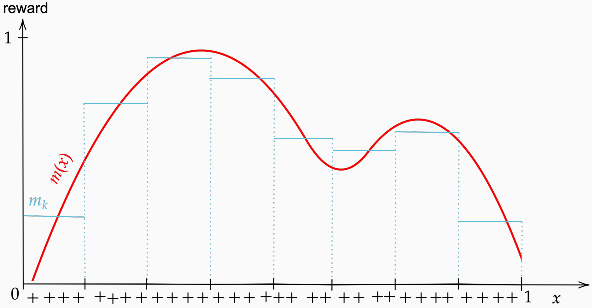
$$|m(x) - m(y)| \leq \max\{|M - m(x)|, L\,|x - y|\}.$$

Assumption (A3) : For all $\epsilon \in (0,1)$,

$$\lambda(\{x : |M - m(x)| \leq \epsilon\}) \leq Q\epsilon.$$

Upper Confidence Bound for Finite bandits

Idea : Discretize the problem by dividing $[0, 1]$ into K intervals, then use UCB on the corresponding finite multi-armed bandit problem.



$m_k = K \int_{I_k} m(a)da.$

Finite Multi-Armed Bandit (FMAB) A player is presented with a set of K actions.

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

- the agent chooses an action $k_t \in \{1, \dots, K\}$;
- she receives the reward y_t such that $\mathbb{E}[y_t | \phi(t)] = m_{k_t}$;
- each action k can be played at most N_k times.

Upper Confidence Bound for Finite bandits :[1]

Parameters : K, δ

Initalization :

- Divide $[0, 1]$ into K intervals
- Choose one action into each interval

For $t = K + 1, ..., T$ **do** :

- Choose interval $k_t \in \text{argmax}_k \, \widehat{m}_k(n_k(t)) + \sqrt{\frac{\log(T/\delta)}{2n_k(t)}}$
- Choose one action uniformly at random from interval k_t, remove it
- If interval k_t is empty, remove it

1. adapted from Auer et al. (2007)

Analysis of the algorithm

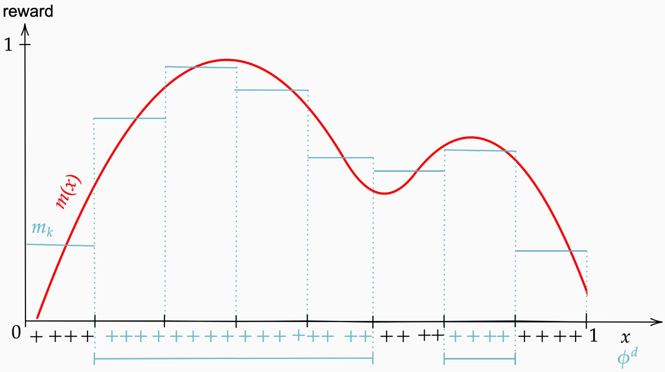
Regret

$$R_T = \sum_{t \leq T} m\left(a_{\phi^*(t)}\right) - \sum_{t \leq T} m\left(a_{\phi(t)}\right).$$

Regret

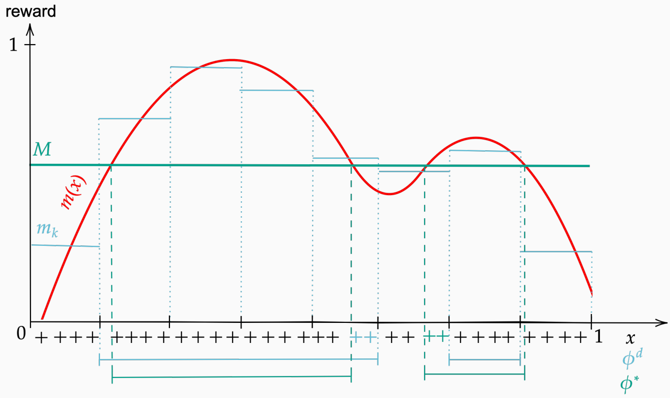
$$\begin{aligned} R_T &= \sum_{t \le T} m\left(a_{\phi^*(t)}\right) - \sum_{t \le T} m\left(a_{\phi(t)}\right) \\ &= \underbrace{\sum_{t \le T} m\left(a_{\phi^*(t)}\right) - \sum_{t \le T} m\left(a_{\phi^d(t)}\right)}_{R_T^{(d)}} + \underbrace{\sum_{t \le T} m\left(a_{\phi^d(t)}\right) - \sum_{t \le T} m\left(a_{\phi(t)}\right)}_{R_T^{(FMAB)}}, \end{aligned}$$

where ϕ^d is the oracle strategy for the discretized problem.



Assume that $m_1 \geq m_2 \geq ... \geq m_k$.

ϕ^d chooses all actions in interval I_1, ..., up to I_f with $f \approx pK$ and $m_f \approx M$.



ϕ^d and ϕ^* mostly select the same actions.

$\Rightarrow R_T^{(d)} = \sum_{t \le T} m\left(a_{\phi^*(t)}\right) - \sum_{t \le T} m\left(a_{\phi^d(t)}\right)$ is small.

Control of $R_T^{(FMAB)}$

$$R_T^{(FMAB)} = \sum_{t \leq T} m\left(a_{\phi^d(t)}\right) - \sum_{t \leq T} m\left(a_{\phi(t)}\right)$$

Control of $R_T^{(FMAB)}$

$$R_T^{(FMAB)} = \sum_{t \leq T} m\left(a_{\phi^d(t)}\right) - \sum_{t \leq T} m\left(a_{\phi(t)}\right)$$

Under (A2), if $a \in I_k$, $m(a) \approx m_k$. Then

$$R_T^{(FMAB)} \approx \sum_{k \leq f} N_k m_k - \sum_{k \leq K} n_k(T) m_k$$

Control of $R_T^{(FMAB)}$

$$R_T^{(FMAB)} = \sum_{t \leq T} m\left(a_{\phi^d(t)}\right) - \sum_{t \leq T} m\left(a_{\phi(t)}\right)$$

Under (A2), if $a \in I_k$, $m(a) \approx m_k$. Then

$$\begin{aligned} R_T^{(FMAB)} &\approx \sum_{k \leq f} N_k m_k - \sum_{k \leq K} n_k(T) m_k \\ &\approx \sum_{k \leq f} (N_k - n_k(T))(m_k - M) + \sum_{k > f} n_k(T)(M - m_k) \end{aligned}$$

$$R_T^{(FMAB)} \approx \sum_{k \leq f} (N_k - n_k(T))(m_k - M) + \sum_{k > f} n_k(T)(M - m_k)$$

- Intervals $k > f$ are sub-optimal : we bound $\sum_{k>f} n_k(T)(M - m_k)$ using classical arguments for continuum-armed bandits.

- Intervals $k \leq f$ are optimal, but with different rewards. We show that all intervals $k \leq f - C$ are exhausted.

Results

Theorem (G. '21)

For the choice $K = N^{1/3} \log(N)^{-2/3}$, *if* $K > p^{-1}$,

$$R_T \leq C_{L,Q} \, (T/p)^{1/3} \log(T/p)^{4/3}$$

with probability $1 - o(1)$.

Remarks :

- Matching lower bounds up to $C_{L,Q} \log(T/p)^{4/3}$;
- In classical continuum-armed bandits, under similar assumptions on m, $K = \sqrt{T}/\log(T)$ and $R_T \leq \sqrt{T}\log(T)$;
  $\Rightarrow$ regrets in the FCAB are lower than in classical CAB!

Conclusion

Finite Continuum-Armed Bandits model situations of sequential allocation of T resources between N competing options.

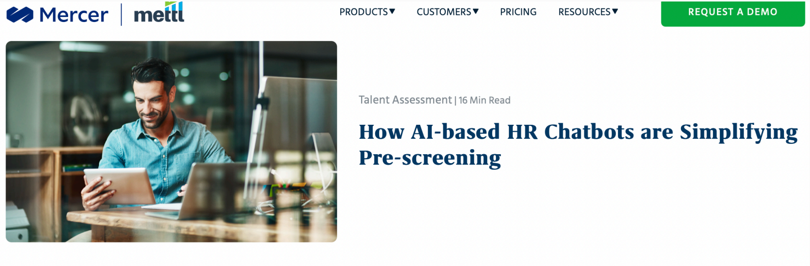
When T/N is fixed, exploration-exploitation trade-off changes : smaller K leads to lower regret rates.

As $T/N \to 0$, regret rate and optimal K increases, and the problem reduces to a classical Continuum-Armed Bandit.

BIASED LINEAR BANDITS

Why fairness in machine learning?

Machine Learning is ubiquitous in daily life.



Why fairness in machine learning?

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.

Linear bandit for sequential decision : At each round $t = 1, \dots, T$

- the agent chooses an action $x_t \in \mathcal{X} \subset \mathbb{R}^d$;
- she receives the reward $x_t^\top \gamma$;
- she observes some feedback $y_t = x_t^\top \gamma + \xi_t$, $\xi_t \overset{i.i.d.}{\sim} \mathcal{N}(0, 1)$.

Goal : Minimize the regret

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

Linear bandits for sequential decision

Linear bandit for sequential decision : At each round $t = 1, \dots, T$

- the agent chooses an action $x_t \in \mathcal{X} \subset \mathbb{R}^d$;
- she receives the reward $x_t^\top \gamma$;
- she observes some feedback $y_t = x_t^\top \gamma + \xi_t$, $\xi_t \overset{i.i.d.}{\sim} \mathcal{N}(0,1)$.

Goal : Minimize the regret

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

Classical algorithms : LINEAR UPPER CONFIDENCE BOUND [2], PHASED ELIMINATION [3], INFORMATION DIRECTED SAMPLING [4]

Assumption : Rewards are bounded : $|x^\top \gamma| \leq 1$, $|\mathcal{X}| < \infty$

Theorem (LATTIMORE and SZEPESVÁRI (2020))

The regret of PHASED ELIMINATION fulfills

$$R_T \leq C\sqrt{dT \log(|\mathcal{X}|T)}.$$

2. ABBASI-YADKORI ET AL. (2011)
3. LATTIMORE and SZEPESVÁRI (2011)
4. KIRSCHNER ET AL. (2020)

Sequential decisions with unfair feedbacks

Biased linear bandit : At each round $t = 1, \dots, T$

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

Assumption : $|\mathcal{X}| < \infty$; $|x^\top \gamma| \leq 1$.

Goal : Minimize the regret

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

Sequential decisions with unfair feedbacks

Biased linear bandit : At each round $t = 1, \dots, T$

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

Assumption : $|\mathcal{X}| < \infty$; $|x^\top \gamma| \leq 1$.

Goal : Minimize the regret

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

Sequential decisions with unfair feedbacks

Biased linear bandit : At each round $t = 1, \dots, T$

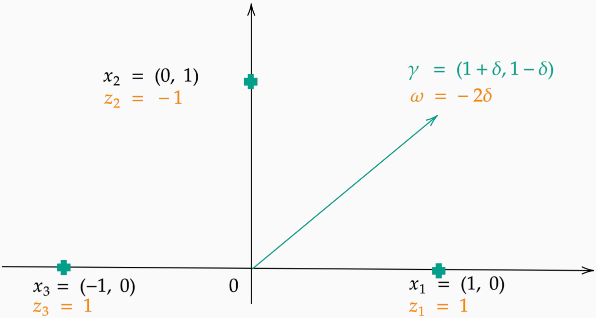
- the agent chooses an action $x_t \in \mathcal{X} \subset \mathbb{R}^d$, described by a sensitive attribute $z_{x_t} \in \{-1, 1\}$;
- she receives the unobserved reward $x_t^\top \gamma$;
- she observes a biased feedback $y_t = x_t^\top \gamma + z_{x_t} \omega + \xi_t$.

Assumption : $|\mathcal{X}| < \infty$; $|x^\top \gamma| \leq 1$.

Goal : Minimize the regret

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

Example



- x_1 is the best action, x_2 is near-optimal, x_3 is very sub-optimal ;
- we need to estimate the bias with precision δ
  $\Rightarrow$ we need to sample x_3 many times ☹.

Outline of Fair Phased Elimination

Main difficulty

It is easy to estimate $a_x^\top \theta$, with $a_x = \binom{x}{z_x}$ and $\theta = \binom{\gamma}{\omega}$, but harder to estimate $x^\top \gamma$.

Main ideas

1. Within a group, feedback and rewards are the same (up to an additive constant) : we can use usual linear bandit technics such as **phased elimination**.
2. To compare action across groups, **estimate the bias independently**.

Bias Estimation

If we sample each action $x \in \mathcal{X}$ exactly $m\mu(x)$ times, the Ordinary Least Square estimator is

$$\hat{\theta} = V^{+}(m\mu) \sum_{t \le n} a_{x_t} y_t, \text{ where } V(m\mu) = \sum_{x \in \mathcal{X}} m\mu(x) a_x a_x^{\top} = mV(\mu).$$

Confidence bound :

Bias Estimation

If we sample each action $x \in \mathcal{X}$ exactly $m\mu(x)$ times, the Ordinary Least Square estimator is

$$\hat{\theta} = V^{+}(m\mu) \sum_{t \leq n} a_{x_t} y_t, \text{ where } V(m\mu) = \sum_{x \in \mathcal{X}} m\mu(x) a_x a_x^\top = mV(\mu).$$

Confidence bound : For all $u \in \mathrm{Range}(V(\mu))$,

$$\mathbb{P}\left(\left| \left(\hat{\theta} - \theta\right)^\top u \right| \leq \sqrt{2m^{-1} \|u\|_{V(\mu)^+}^2 \log\left(\frac{1}{\delta}\right)} \right) \geq 1 - \delta.$$

where $\|u\|_{V(\mu)^+}^2 := u^\top V(\mu)^+ u.$

Bias Estimation

If we sample each action $x \in \mathcal{X}$ exactly $m\mu(x)$ times, the Ordinary Least Square estimator is

$$\hat{\theta} = V^{+}(m\mu)\sum_{t \le n} a_{x_t} y_t, \text{ where } V(m\mu) = \sum_{x \in \mathcal{X}} m\mu(x) a_x a_x^\top = mV(\mu).$$

Confidence bound : If $e_{d+1} \in \mathrm{Range}(V(\mu))$,

$$\mathbb{P}\left(|\hat{\omega} - \omega| \le \sqrt{2m^{-1} \left\| e_{d+1} \right\|^2_{V(\mu)^+} \log\left(\frac{1}{\delta}\right)} \right) \ge 1 - \delta.$$

since $\omega = \theta^\top e_{d+1}$.

First idea : Estimate $\omega = \theta^\top e_{d+1}$ using $\mathbf{e_{d+1}}$**-optimal design**

Find $$\mu^* \in \operatorname*{argmin}_{\mu \in \mathcal{P}_{e_{d+1}}} \|e_{d+1}\|^2_{V(\mu)^+}.$$

Set $$\kappa_* = \|e_{d+1}\|^2_{V(\mu^*)^+}.$$

First idea : Estimate $\omega = \theta^\top e_{d+1}$ using $\mathbf{e_{d+1}}$**-optimal design**

Find $$\mu^* \in \underset{\mu \in \mathcal{P}_{e_{d+1}}}{\operatorname{argmin}} \|e_{d+1}\|^2_{V(\mu)^+}.$$

Set $$\kappa_* = \|e_{d+1}\|^2_{V(\mu^*)^+}.$$

Second idea : Estimate $\omega = \theta^\top e_{d+1}$ using **Δ-optimal design**

For $\Delta_x = (x^* - x)^\top \gamma$, $\Delta = (\Delta_x)_{x \in \mathcal{X}}$, find

$$\mu^\Delta \in \operatorname*{argmin}_{\mu \in \mathcal{M}_{e_{d+1}}} \sum_{x \in \mathcal{X}} \mu(x) \Delta_x \quad \text{such that } \|e_{d+1}\|^2_{V(\mu)^+} = 1.$$

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

Second idea : Estimate $\omega = \theta^\top e_{d+1}$ using **Δ-optimal design**

For $\Delta_x = (x^* - x)^\top \gamma$, $\Delta = (\Delta_x)_{x \in \mathcal{X}}$, find

$$\mu^\Delta \in \underset{\mu \in \mathcal{M}_{e_{d+1}}}{\operatorname{argmin}} \sum_{x \in \mathcal{X}} \mu(x) \Delta_x \quad \text{such that } \|e_{d+1}\|^2_{V(\mu)^+} = 1.$$

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

Theorem (G., Carpentier, Giraud (2022))

Fair Phased Elimination algorithm fulfills

$$R_T \leq C\kappa_*^{1/3} \log(T)^{1/3} T^{2/3}$$

for large T.

Remarks

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

Worst-case regret

Theorem (G., Carpentier, Giraud (2022))

Fair Phased Elimination algorithm fulfills

$$R_T \leq C\kappa_*^{1/3} \log(T)^{1/3} T^{2/3}$$

for large T.

Remarks

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

Gap-dependent regret

Theorem (G., Carpentier, Giraud (2022))

Fair Phased Elimination algorithm fulfills

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

where $\Delta_{\min} = \min_{x \neq x^*} \Delta_x$, $\Delta_{\neq} = \min_{z_x \neq z_{x^*}} \Delta_x$ *and* $\varepsilon_T = \left({}^{\kappa_* \log(T)}/_{T}\right)^{1/3}$.

Remarks

- Matching lower bounds up to numerical constant.
- $\frac{d \log(T)}{\Delta_{\min}}$ 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 feedbacks;

Gap-dependent regret

Theorem (G., Carpentier, Giraud (2022))

Fair Phased Elimination algorithm fulfills

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

where $\Delta_{\min} = \min_{x \neq x^*} \Delta_x$, $\Delta_{\neq} = \min_{z_x \neq z_{x^*}} \Delta_x$ *and* $\varepsilon_T = \left(\kappa_* \log(T)/T\right)^{1/3}$.

Remarks

- Matching lower bounds up to numerical constant.
- $\frac{d \log(T)}{\Delta_{\min}}$ 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 feedbacks;

Theorem (G., Carpentier, Giraud (2022))

Fair Phased Elimination algorithm fulfills

$$R_T \leq C \left(\frac{d}{\Delta_{\min}} \vee \frac{\kappa(\Delta \vee \Delta_{\neq} \vee \varepsilon_T)}{\Delta_{\neq}^2} \right) \log(T) \qquad \text{for large } T,$$

where $\Delta_{\min} = \min_{x \neq x^*} \Delta_x$, $\Delta_{\neq} = \min_{z_x \neq z_{x^*}} \Delta_x$ *and* $\varepsilon_T = \left(\kappa_* \log(T) / T \right)^{1/3}$.

Remarks

- Matching lower bounds up to numerical constant.
- $\frac{d \log(T)}{\Delta_{\min}}$ 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 feedbacks;

Conclusion

- Biased linear bandits model sequential decision-making scenarii with biased observations.
- 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 margin to a separating hyperplane.
- 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 $\frac{\kappa(\Delta)\log(T)}{\Delta_{\neq}^2} \leq \frac{d\log(T)}{\Delta_{\min}}$.

Thank you for your attention!

Bibliography

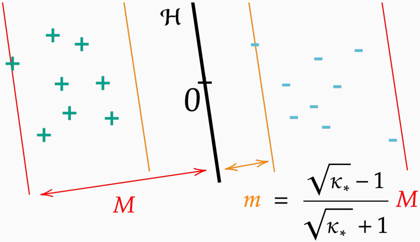
- Auer et al. (2007). Improved Rates for the Stochastic Continuum-Armed Bandit Problem. *COLT'07*.
- Abbasi-Yadkori et al. (2011). Improved Algorithms for Linear Stochastic Bandits. *NIPS'11*.
- Bubeck et al. (2011). X-Armed Bandits. *Journal of Machine Learning Research*.
- Gao et al. (2015). Rate-optimal graphon estimation. *EJS*.

Bibliography

- Gao et al. (2016). Optimal Estimation and Completion of Matrices with Biclustering Structures. *JMLR*.
- Kirschner et al. (2020). Information Directed Sampling for Linear Partial Monitoring. *COLT'20*.
- Kleinberg, Robert (2004). Nearly Tight Bounds for the Continuum-Armed Bandit Problem. *NIPS'04*.
- Klopp et al. (2017). Oracle inequalities for network models and sparse graphon estimation. *AOS*.
- Lattimore and Szepesvári (2020). Bandit Algorithms. *Cambridge University Press*.

Appendix

Geometry of bias estimation



Lemma

κ_* is the largest $\kappa \geq 0$ such that there exists an hyperplane $\mathcal{H}$ separating the two groups with $m = \frac{\sqrt{\kappa}-1}{\sqrt{\kappa}+1} M$, where :

- *m is the margin to $\mathcal{H}$;*
- *M is the maximum distance of all points to the hyperplane.*

FCAB WITH $T/N \to 0$

Corollary (G. (2021))

Assume that $T = 0.5N^{\alpha}$ *for some* $\alpha \in (2/3 + \epsilon_N{}^{5}, 1]$. *For the choice* $K = \alpha^{-2/3}(2T)^{1/(2\alpha)} \log(2T)^{-2/3}$ *and* $\delta = N^{-4/3}$,

$$R_T \leq C_{Q,L} T^{1/(3\alpha)} \log(T)^{4/3}$$

with probability $1 - o(1)$.

Remarks

- When $p \to 0$, regret increases from FCAB to CAB regime.
- $\alpha = 2/3 + \epsilon_N$ corresponds to transition from FCAB to CAB. Then, $T = N/K$:
  - all optimal actions are in 1 interval;
  - no interval is ever exhausted.

5. $\epsilon_N = \left(\frac{2}{3} \log\log(N) + \log(2)\right) / \log(N)$

SBM approximation of smooth graphon

Assume full observation : $\Pi_{i,j} = 1$.

α-Hoelder regularity assumption : For all $(x, y) \in [0,1]^2$,

$$\left|W(x', y') - \mathcal{P}_{\lfloor\alpha\rfloor}{}^{6}((x,y),(x'-x, y'-y))\right| \leq M\left(|x-x'|^{\alpha-\lfloor\alpha\rfloor} + |y-y'|^{\alpha-\lfloor\alpha\rfloor}\right)$$

Corollary (G., Klopp (2021))

If $\rho_n > n^{-1}$, for $k = \left\lceil n^{\frac{1}{1+(\alpha\wedge1)}} \rho_n^{\frac{1}{2+2(\alpha\wedge1)}} \right\rceil$,

$$\|\Theta^* - \widehat{\Theta}\|_F^2 \leq C_c \rho_n \left(n^{\frac{2}{1+(\alpha\wedge1)}} \rho_n^{\frac{1}{1+(\alpha\wedge1)}} + n \log(\rho_n n) \right)$$

with probability at least $1 - 9\exp(-C\rho_n n \log(k))$.

6. $\mathcal{P}_{\lfloor\alpha\rfloor}((x,y),\cdot)$ is the Taylor polynomial of W of degree $\lfloor\alpha\rfloor$ at point (x,y)