Structural Causal Bandits with non-manipulable variables

Sanghack Lee Elias Bareinboim

Purdue University

Executive Summary

- **SCM-MAB** = MAB + Causality

Executive Summary

- **SCM-MAB** = MAB + Causality

  actions = interventions

Executive Summary

- **SCM-MAB** = MAB + Causality
  actions = interventions
- Q: Which interventions *can* be optimal?

Executive Summary

- **SCM-MAB** = MAB + Causality
  actions = interventions

- Q: Which interventions *can* be optimal?
  A: Possibly-Optimal Minimal Intervention Set (**POMIS**)

Executive Summary

- **SCM-MAB** = MAB + Causality
  actions = interventions
- Q: Which interventions *can* be optimal?
  A: Possibly-Optimal Minimal Intervention Set (**POMIS**)
- Q: How to learn *an arm*'s reward from *other arms*?

Executive Summary

- **SCM-MAB** = MAB + Causality
  actions = interventions
- Q: Which interventions *can* be optimal?
  A: Possibly-Optimal Minimal Intervention Set (**POMIS**)
- Q: How to learn *an arm*'s reward from *other arms*?
  A: a generalized identifiability algorithm (**z^{2}ID**)

Executive Summary

- **SCM-MAB** = MAB + Causality
  actions = interventions
- Q: Which interventions *can* be optimal?
  A: Possibly-Optimal Minimal Intervention Set (**POMIS**)
- Q: How to learn *an arm*'s reward from *other arms*?
  A: a generalized identifiability algorithm ($\mathbf{z^2}$**ID**)
- Faster convergence: smaller arms to play; more accurate estimation.

Motivation

Multi-armed bandit (MAB)

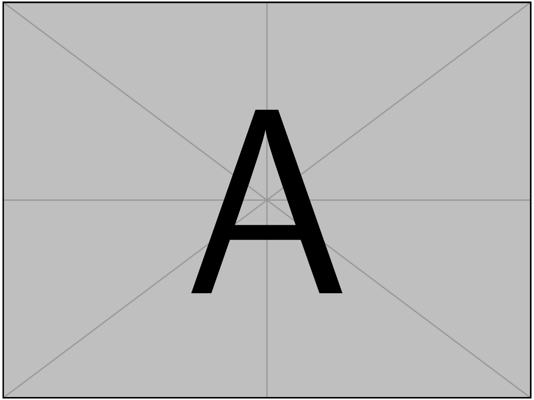
A classic, sequential decision-making problem

- Given: a set of **arms** (actions), $\mathbf{A}$
- How: at round t, pull an arm A_t, and get a **reward** Y_{A_t}
- Goal: to minimize cumulative **regret** (or maximize cumulative reward)

- a trade-off between **exploitation** vs. **exploration**
- Examples: ad. placement, medical treatment, packet routing ...
- Key **assumption**: arms are *independent* (in a traditional MAB setting).

Multi-armed bandit (MAB)

The action & reward can be understood as,



Multi-armed bandit through Causal Lens

Why?

- environments can be modeled explicitly with causal graphs

decision = passive? (observation) active? (intervention, where? how?) not every variable is intervenable
(figure, intervention, causal mechanism, graphs, outcome) ((what’s causal about MAB?))

Multi-armed bandit through Causal Lens

Structural Causal Model (SCM) $\mathcal{M} = \langle \mathbf{U}, \mathbf{V}, \mathbf{F}, P(\mathbf{U}) \rangle$:

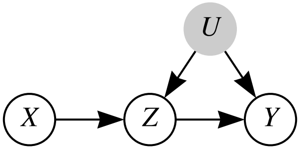
- $\mathbf{U}$: unobserved variables
- $\mathbf{V}$: observed variables
- $\mathbf{F}$: a set of functions for $\mathbf{V}$
- $P(\mathbf{U})$: a joint distribution over $\mathbf{U}$ ($\sim$randomness)

Multi-armed bandit through Causal Lens

Structural Causal Model (SCM) $\mathcal{M} = \langle \mathbf{U}, \mathbf{V}, \mathbf{F}, P(\mathbf{U}) \rangle$:

- $\mathbf{U}$: unobserved variables
- $\mathbf{V}$: observed variables
- $\mathbf{F}$: a set of functions for $\mathbf{V}$
- $P(\mathbf{U})$: a joint distribution over $\mathbf{U}$ ($\sim$randomness)

A causal graph $\mathcal{G}$ conforming to $\mathcal{M}$ looks like a **DAG** + **bidirected edges** for **unobserved confounders** (UCs).[1]



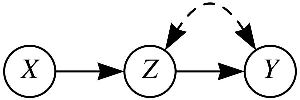
[1] among $\mathbf{U}$, only UCs will be visualized.

Multi-armed bandit through Causal Lens

Structural Causal Model (SCM) $\mathcal{M} = \langle \mathbf{U}, \mathbf{V}, \mathbf{F}, P(\mathbf{U}) \rangle$:

- $\mathbf{U}$: unobserved variables
- $\mathbf{V}$: observed variables
- $\mathbf{F}$: a set of functions for $\mathbf{V}$
- $P(\mathbf{U})$: a joint distribution over $\mathbf{U}$ ($\sim$randomness)

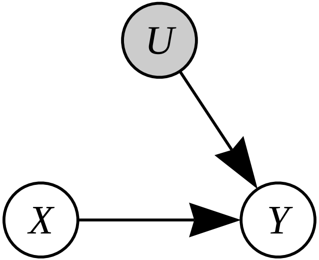
A causal graph $\mathcal{G}$ conforming to $\mathcal{M}$ looks like a **DAG** + **bidirected edges** for **unobserved confounders** (UCs).[1]



[1] among $\mathbf{U}$, only UCs will be visualized.

Multi-armed bandit through Causal Lens

An example with a traditional MAB problem



(problem with do nothing?)

SCM-MAB

SCM-MAB

A **SCM-MAB** is $\langle M, Y, \mathbf{N} \rangle$:

- a SCM $\mathcal{M} = \langle \mathbf{U}, \mathbf{V}, \mathbf{F}, P(\mathbf{U}) \rangle$,
- a reward variable $Y \in \mathbf{V}$,
- non-manipulable variables $\mathbf{N} \subseteq \mathbf{V} \setminus \{Y\}$

SCM-MAB

A **SCM-MAB** is $\langle M, Y, \mathbf{N} \rangle$:

- a SCM $\mathcal{M} = \langle \mathbf{U}, \mathbf{V}, \mathbf{F}, P(\mathbf{U}) \rangle$,
- a reward variable $Y \in \mathbf{V}$,
- non-manipulable variables $\mathbf{N} \subseteq \mathbf{V} \setminus \{Y\}$

Therefore,

- Actions: $\mathbf{A} = \{\mathbf{x} \in \mathfrak{X}_{\mathbf{X}} | \mathbf{X} \subseteq \mathbf{V} \setminus \mathbf{N} \setminus \{Y\}\}$ (including observation)
- Reward distribution: $P(Y|do(\mathbf{X} = \mathbf{x}))$ (or $P_{\mathbf{x}}(Y)$) $(\forall_{\mathbf{x} \in \mathbf{A}})$
- Expected reward: $\mu_{\mathbf{x}} = \mathbb{E}[Y|do(\mathbf{x})]$

SCM-MAB

A **SCM-MAB** is $\langle M, Y, \mathbf{N} \rangle$:

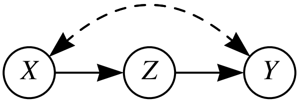
- a SCM $\mathcal{M} = \langle \mathbf{U}, \mathbf{V}, \mathbf{F}, P(\mathbf{U}) \rangle$,
- a reward variable $Y \in \mathbf{V}$,
- non-manipulable variables $\mathbf{N} \subseteq \mathbf{V} \setminus \{Y\}$

Therefore,

- Actions: $\mathbf{A} = \{\mathbf{x} \in \mathfrak{X}_{\mathbf{X}} | \mathbf{X} \subseteq \mathbf{V} \setminus \mathbf{N} \setminus \{Y\}\}$ (including observation)
- Reward distribution: $P(Y|do(\mathbf{X} = \mathbf{x}))$ (or $P_{\mathbf{x}}(Y)$) $(\forall_{\mathbf{x} \in \mathbf{A}})$
- Expected reward: $\mu_{\mathbf{x}} = \mathbb{E}[Y|do(\mathbf{x})]$

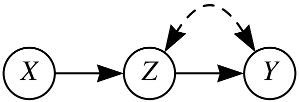
Assumption: we access to the causal graph $\mathcal{G}$ without knowing $\mathbf{F}$ nor $P(\mathbf{U})$.

SCM-MAB examples



- Intervention sets: $\emptyset$, $\{X\}$, $\{Z\}$, $\{X, Z\}$
- Arms: There are 9 arms: $do(\emptyset)$, $do(X = 0)$, $do(X = 1)$, $\ldots$, $do(X = 1, Z = 1)$

SCM-MAB examples



- Intervention sets: $\emptyset$, $\{X\}$, $\{Z\}$, $\{X, Z\}$
- Arms: $do(\emptyset)$, $do(X=0)$, $do(X=1)$, $\ldots$, $do(X=1, Z=1)$

Structural Properties

Arms are **dependent** through underlying causal mechanism in SCM-MAB.

1. **Equivalence**: two arms share the **same** reward distribution

Structural Properties

Arms are **dependent** through underlying causal mechanism in SCM-MAB.

1. **Equivalence**: two arms share the **same** reward distribution

$$\mu_{\mathbf{x}} = \mu_{\mathbf{z}}$$

Structural Properties

Arms are **dependent** through underlying causal mechanism in SCM-MAB.

1. **Equivalence**: two arms share the **same** reward distribution

$$\mu_{\mathbf{x}} = \mu_{\mathbf{z}}$$

2. **Partial-orders**: one arm is always preferred to the other

Structural Properties

Arms are **dependent** through underlying causal mechanism in SCM-MAB.

1. **Equivalence**: two arms share the **same** reward distribution

$$\mu_{\mathbf{x}} = \mu_{\mathbf{z}}$$

2. **Partial-orders**: one arm is always preferred to the other

$$\mu_{\mathbf{x}^*} \geq \mu_{\mathbf{z}^*}$$

Structural Properties

Arms are **dependent** through underlying causal mechanism in SCM-MAB.

1. **Equivalence**: two arms share the **same** reward distribution

$$\mu_{\mathbf{x}} = \mu_{\mathbf{z}}$$

2. **Partial-orders**: one arm is always preferred to the other

$$\mu_{\mathbf{x}^*} \geq \mu_{\mathbf{z}^*}$$

3. **Expressions**: infer one arm's reward distr. from other arms' samples.

Structural Properties

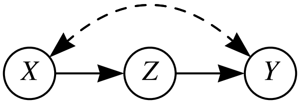
Arms are **dependent** through underlying causal mechanism in SCM-MAB.

1. **Equivalence**: two arms share the **same** reward distribution
$$\mu_{\mathbf{x}} = \mu_{\mathbf{z}}$$
2. **Partial-orders**: one arm is always preferred to the other
$$\mu_{\mathbf{x}^*} \geq \mu_{\mathbf{z}^*}$$
3. **Expressions**: infer one arm's reward distr. from other arms' samples.
$$P_x(y) = \textstyle\sum_z P(z|x) \sum_{x'} P(y|z, x')P(x')$$

Structural Properties in SCM-MAB

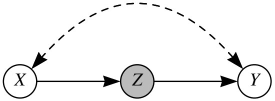
Arms are **dependent** through underlying causal mechanism in SCM-MAB.

1. **Equivalence**: two arms share the **same** reward distribution
$$\mu_{x,z} = \mu_z$$
2. **Partial-orders**: one arm is always preferred to the other
$$\mu_{x^*} \geq \mu_{z^*}$$
3. **Expressions**: infer one arm's reward distr. from other arms' samples.
$$P_x(y) = \sum_z P(z|x) \sum_{x'} P(y|z, x')P(x')$$



Structural Property 1: Equivalence

Consider a graph, named *front-door*:



Definition (Minimal Intervention Set (MIS))

Given $\langle \mathcal{G}, Y, \mathbf{N} \rangle$, a set of variables $\mathbf{X} \subseteq \mathbf{V} \setminus \{Y\} \setminus \mathbf{N}$ is said to be a *minimal intervention set* if there is no $\mathbf{X}' \subset \mathbf{X}$ such that $\mu_{\mathbf{x}'} = \mu_{\mathbf{x}}$ for every SCM conforming to $\mathcal{G}$ where $\mathbf{x}' \in \mathfrak{X}_{\mathbf{X}'}$ that is consistent with $\mathbf{x}$.

Structural Property 2: Partial-orderedness

Structural Property 2: Partial-orderedness

Possibly-Optimal Minimal Intervention Set
Key theorem without non-manipulability constraints
(soundness-and-completeness theorem goes here)

Structural Property 2: Partial-orderedness

Definition (Possibly-Optimal Minimal Intervention Set (POMIS))

Given $\langle \mathcal{G}, Y, \mathbf{N} \rangle$, let $\mathbf{X} \in \mathbb{M}^{\mathbf{N}}_{\mathcal{G},Y}$. If there exists a SCM conforming to $\mathcal{G}$ such that $\mu_{\mathbf{x}^*} > \forall_{\mathbf{W} \in \mathbb{M}^{\mathbf{N}}_{\mathcal{G},Y} \setminus \{\mathbf{X}\}} \mu_{\mathbf{w}^*}$, then $\mathbf{X}$ is a *possibly-optimal minimal intervention set* with respect to $\langle \mathcal{G}, Y, \mathbf{N} \rangle$.

Structural Property 2: Partial-orderedness

Definition (Possibly-Optimal Minimal Intervention Set (POMIS))

Given $\langle \mathcal{G}, Y, \mathbf{N} \rangle$, let $\mathbf{X} \in \mathbb{M}^{\mathbf{N}}_{\mathcal{G},Y}$. If there exists a SCM conforming to $\mathcal{G}$ such that $\mu_{\mathbf{x}^*} > \forall_{\mathbf{W} \in \mathbb{M}^{\mathbf{N}}_{\mathcal{G},Y} \setminus \{\mathbf{X}\}} \mu_{\mathbf{w}^*}$, then $\mathbf{X}$ is a *possibly-optimal minimal intervention set* with respect to $\langle \mathcal{G}, Y, \mathbf{N} \rangle$.

Key theorem with non-manipulability constraints Latent projection here?

Structural Property 3: Relating (POMISs) Arms

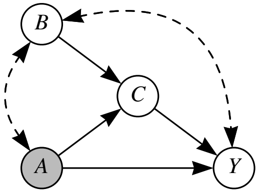
- Q: how are samples from $\{do(\mathbf{z})\}_{\mathbf{Z} \in \text{POMIS}}$ related to $do(\mathbf{x})$?

Structural Property 3: Relating (POMISs) Arms

- Q: how can we express $P_{\mathbf{x}}(\mathbf{v}')$ with $\{P_{\mathbf{z}}\}_{\mathbf{Z} \in \text{POMIS}}$?

- ID: $P_{\mathbf{x}}(\mathbf{v}')$ from $P(\mathbf{v})$ (SP, 2006)
- zID: $P_{\mathbf{x}}(\mathbf{v}')$ from $P_{\mathbf{z}'}(\mathbf{v})$ for $\mathbf{Z}' \subseteq \mathbf{Z}$ (BP, 2012)
- **z^{2}ID**: $P_{\mathbf{x}}(\mathbf{v}')$ from a set of experiments (LB, 2019)

Structural Property 3: Relating POMISs — an example



POMISs are $\emptyset$, $\{B\}$, and $\{C\}$.

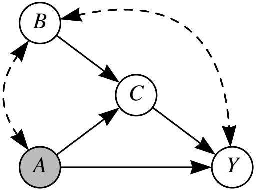
$$P(y) = \sum_{a,b,c} P_b(c|a)P_c(a,b,y)$$

$$P_b(y) = \sum_{a,c} P(c|a,b) \sum_{b'} P(y|a,b',c)P(a,b')$$

$$P_c(y) = \sum_{a,b} P(y|a,b,c)P(a,b)$$

$$P_c(y) = \sum_{a} P_b(y|a,c)P_b(a)$$

Structural Property 3: Relating POMISs — an example



POMISs are $\emptyset$, $\{B\}$, and $\{C\}$.

$$P(y) = \sum_{a,b,c} P_b(c|a) P_c(a,b,y)$$

$$P_b(y) = \sum_{a,c} P(c|a,b) \sum_{b'} P(y|a,b',c) P(a,b')$$

$$P_c(y) = \sum_{a,b} P(y|a,b,c) P(a,b)$$

$$P_c(y) = \sum_{a} P_b(y|a,c) P_b(a)$$

SCM-MAB algorithm

SCM-MAB algo.: MAB algo. taking advantage of the structural properties.

function z²-TS$(\mathcal{G}, Y, \mathbf{N}, T)$

$\mathbb{Z} \leftarrow \mathbb{P}^{\mathbf{N}}_{\mathcal{G},Y}$

$\mathbf{A} \leftarrow \{\mathbf{x} \in \mathfrak{X}_{\mathbf{X}} \mid \mathbf{X} \in \mathbb{Z}\}$

$\hat{\boldsymbol{\theta}}_{\mathbf{x}} \leftarrow \{P_{\mathbf{x}}(y)\} \cup \{\text{z}^2\text{ID}(\mathcal{G}, y, \mathbf{x}, \mathbb{Z}')\}_{\mathbb{Z}' \subseteq \mathbb{Z} \setminus \{\mathbf{X}\}}$ **for** $\mathbf{x} \in \mathbf{A}$

$\mathbf{D} \leftarrow \{D_{\mathbf{x}} = \emptyset\}_{\mathbf{x} \in \mathbf{A}}$

for t in $1, \ldots, T$ **do**

for $\mathbf{x} \in \mathbf{A}$ **do**

$\hat{\theta}_{\mathbf{x}}, \hat{s}^2_{\mathbf{x}} \leftarrow \text{bMVWA}(\mathbf{D}, \hat{\boldsymbol{\theta}}_{\mathbf{x}})$

Find $\hat{\alpha}_{\mathbf{x}}, \hat{\beta}_{\mathbf{x}}$ such that $\text{Beta}(\hat{\alpha}_{\mathbf{x}}, \hat{\beta}_{\mathbf{x}})$ matching $\hat{\theta}_{\mathbf{x}}, \hat{s}^2_{\mathbf{x}}$

$\theta_{\mathbf{x}} \sim \text{Beta}(\hat{\alpha}_{\mathbf{x}}, \hat{\beta}_{\mathbf{x}})$

$\mathbf{x}' \leftarrow \arg\max_{\mathbf{x} \in \mathbf{A}} \theta_{\mathbf{x}}$

Sample $\mathbf{v}$ by $do(\mathbf{x}')$ and append $\mathbf{v}$ to $D_{\mathbf{x}'}$

SCM-MAB algorithm

SCM-MAB algo.: MAB algo. taking advantage of the structural properties.

function z^2-KL-UCB$(\mathcal{G}, Y, \mathbf{N}, T, f \leftarrow \ln(t)+3\ln(\ln(t)))$

Initialize $\mathbb{Z}$, $\mathbf{A}$, $\{\hat{\boldsymbol{\theta}}_{\mathbf{x}}\}_{\mathbf{x}\in\mathbf{A}}$, $\mathbf{D}$

$(\forall_{\mathbf{x}\in\mathbf{A}})$ Sample $\mathbf{v}$ by $do(\mathbf{x})$, and append $\mathbf{v}$ to $D_{\mathbf{x}}$

for t in $|\mathbf{A}|, \ldots, T$ **do**

$\hat{\theta}_{\mathbf{x}}, \hat{s}^2_{\mathbf{x}} \leftarrow$ bMVWA$(\mathbf{D}, \hat{\boldsymbol{\theta}}_{\mathbf{x}})$ **for** $\mathbf{x} \in \mathbf{A}$

$\hat{N}_{\mathbf{x}} \leftarrow \hat{\theta}_{\mathbf{x}}(1-\hat{\theta}_{\mathbf{x}})/\hat{s}^2_{\mathbf{x}}$; $\quad \hat{t} \leftarrow \sum_{\mathbf{x}} \hat{N}_{\mathbf{x}}$

$$\boldsymbol{\mu} = \left\{ \sup \left\{ \mu \in [0,1] : \mathrm{KL}(\hat{\theta}_{\mathbf{x}}, \mu) \leq \frac{f(\hat{t})}{\hat{N}_{\mathbf{x}}} \right\} \right\}_{\mathbf{x}\in\mathbf{A}}$$

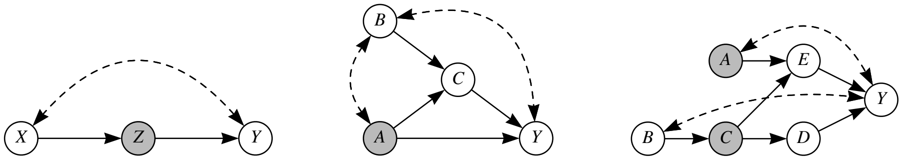
$\mathbf{x}' \leftarrow \arg\max_{\mathbf{x}\in\mathbf{A}} \mu_{\mathbf{x}}$

Sample $\mathbf{v}$ by $do(\mathbf{x}')$, and append $\mathbf{v}$ to $D_{\mathbf{x}'}$

Empirical Evaluation

Experimental settings

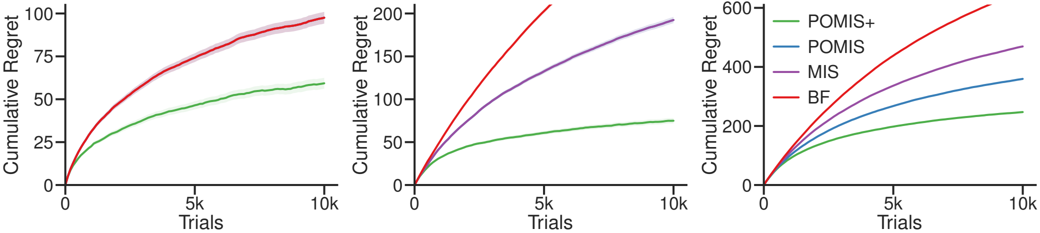
- **4** strategies: Brute-force, MIS, POMIS, POMIS+
- **2** base MAB algorithms: Thompson sampling (TS), kl-UCB
- **3** SCM-MAB problems, binary $\mathbf{V}$



- **1000** simulations

Experimental results (average cumulative regret)

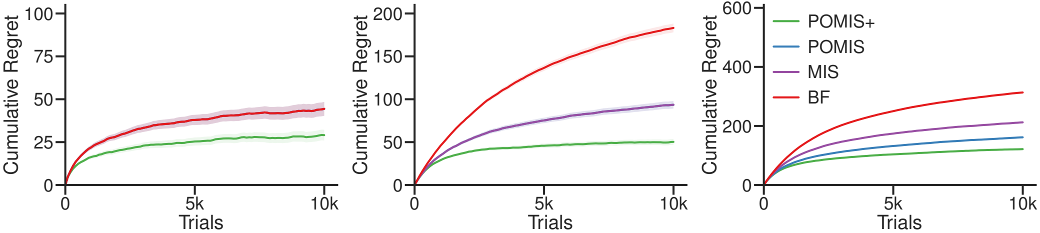
base algorithm: kl-UCB



Performance: POMIS+ $>$ POMIS $\geq$ MIS $\geq$ Brute-force

Experimental results (average cumulative regret)

base algorithm: TS

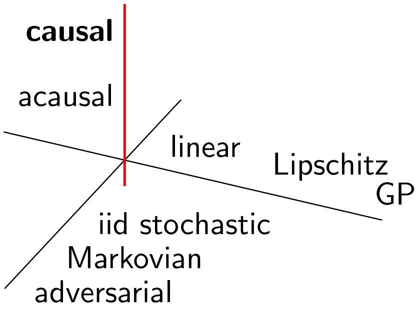


Performance: POMIS+ $>$ POMIS $\geq$ MIS $\geq$ Brute-force

Conclusions

Bandit Landscape

Dimensions: functional assumptions, bandit type, reward type, etc.
We generalized MABs into a causal dimension.



SCM-MAB: stochastic iid reward, nonparametric, bandit feedback

Conclusions

We studied how to take advantage of a known causal graph in bandit setting:

- introduced: **SCM-MAB** w/ non-manipulability constraints
- characterized: Possibly-Optimal Minimal Intervention Set (POMIS)
- devised: z^2ID for connecting arms
- provided: SCM-MAB algorithms: z^2-TS, z^2-kl-UCB