Structural Causal Bandits: Where to Intervene?

Sanghack Lee and Elias Bareinboim `{lee2995,eb}@purdue.edu`
CausalAI Lab, Purdue University

Overview

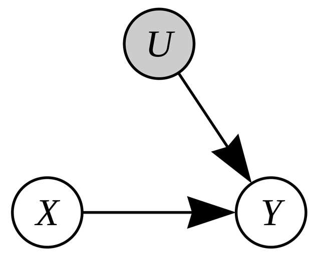
We propose **SCM-MAB**, marrying Multi-armed Bandit (**MAB**) with Structural Causal Model (**SCM**). Whenever the underlying causal mechanism for arms' rewards is well-understood, an agent can play a bandit *more effectively*, while a naive agent ignorant to such a mechanism may be *fail* or *slow* to converge.

Multi-armed bandit (MAB) is one of the prototypical sequential decision-making settings found in various real-world applications.

- **Arms**: There are arms **A** in the bandit (i.e., slot machine); each arm associates with a reward distribution,
- **Play**: an agent plays the bandit by pulling an arm $A_{\mathbf{x}} \in \mathbf{A}$ each round,
- **Reward**: a reward $Y_{\mathbf{x}}$ is drawn from the arm's reward distribution,
- **Goal**: to minimize a cumulative regret (CR) over time horizon T.

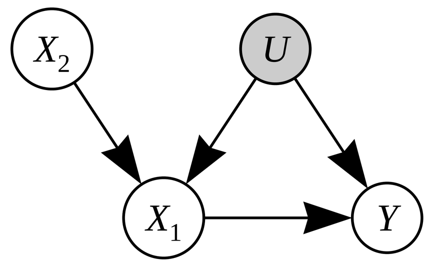
Multi-armed Bandit through Causal Lens

- pulling an arm = intervening on a set of variables (intervention set, IS)
- reward mechanism = causal mechanism



- Formally, playing an arm A_x is setting X to x (called *do*), and observing Y drawn from $P(Y|do(X=x))$ where $P(y|do(x)) := \sum_u \mathbf{1}_{f(x,u),y} P(u)$.

Why do we need Causal MABs? A Motivating Example



- **Q**: How many **arms** are there? (We can control 2 binary variables, X_1 and X_2)
  A: **Nine**. We need to choose a set among
  $$\{\emptyset, \{X_1\}, \{X_2\}, \{X_1, X_2\}\}$$
  and then make the corresponding assignment (**all-subsets**). A *naive* combinatorial agent will intervene on $\{X_1, X_2\}$, simultaneously (= 4 arms).
- **Q**: Why is playing $\{X_1, X_2\}$ (**all-at-once**) considered *naive*?
  A: This strategy may *miss* the optimal arm, as shown in the simulation below:

There exists a environment (i.e., parametrization) where intervening on X_2 is optimal, and intervening on $\{X_1, X_2\}$, simultaneously is always sub-optimal.
e.g., $X_1 = X_2 \oplus U$, $Y = X_1 \oplus U$. (when $X_2=1$, X_1 carries $\neg U$, and Y checks $X_1 \neq U$)

- **Q**: What are the arms **worth** playing, regardless of the parametrization?
  A: Intervening on either $\{X_2\}$ or $\{X_1\}$ can be shown to be sufficient since:
  $\because$ (i) $\max \mu_{x_2} \geq \max \mu_{\emptyset}$, (ii) $\max \mu_{x_1} = \max \mu_{x_1,x_2}$, (iii) $\max \mu_{x_2} <> \max \mu_{x_1}$

SCM-MAB — Connecting Bandits With Structural Causal Models

A Structural Causal Model (**SCM**) $\mathcal{M}$ is a 4-tuple $\langle \mathbf{U}, \mathbf{V}, \mathbf{F}, P(\mathbf{U}) \rangle$:

- **U** is a set of **unobserved** variables (**unknown**);
- **V** is a set of **observed** variables (**known**);
- **F** is a set of **causal mechanisms** for **V** using **U** and **V**;
- $P(\mathbf{U})$ is a joint distribution over the **U** (**randomness**).

The **SCM** allows one to model the underlying causal relations (usually unobserved). The environment where the MAB solver will perform experiments can be modeled as an **SCM**, following the connection established next.

SCM-MAB

- SCM $\mathcal{M} = \langle \mathbf{U}, \mathbf{V}, \mathbf{F}, P(\mathbf{U}) \rangle$ and a reward variable $Y \in \mathbf{V}$, $\langle \mathcal{M}, Y \rangle$
- Arms **A** correspond to *all* interventions $\{A_{\mathbf{x}} | \mathbf{x} \in D(\mathbf{X}), \mathbf{X} \subseteq \mathbf{V} \setminus \{Y\}\}$.
- Reward: distribution $P(Y_{\mathbf{x}}) := P(Y|do(\mathbf{X}=\mathbf{x}))$, expected, $\mu_{\mathbf{x}} := \mathbb{E}[Y|do(\mathbf{X}=\mathbf{x})]$.

We assume that a causal graph $\mathcal{G}$ of $\mathcal{M}$ is accessible, but not $\mathcal{M}$ itself.

SCM-MAB Properties — Dependence Structure Across Arms

1. *Equivalence* among Arms
Two arms share the same reward distribution, i.e.,
$$\mu_{\mathbf{x},\mathbf{z}} = \mu_{\mathbf{x}}$$
whenever intervening on some variables doesn't have a causal effect on the outcome.
$\rightarrow$ Test $P(y|do(\mathbf{x},\mathbf{z})) = P(y|do(\mathbf{x}))$ through $Y \perp\!\!\!\perp \mathbf{Z} \mid \mathbf{X}$ in $\mathcal{G}_{\overline{\mathbf{X} \cup \mathbf{Z}}}$ (*do*-calculus).

— **Minimal Intervention Set (MIS, Def. 1)**

- A **minimal** set of variables among ISs sharing the same reward distribution.
- Given that there are sets with the same reward distribution, we would like to intervene on a *minimal* set of variables yielding smaller # of arms.

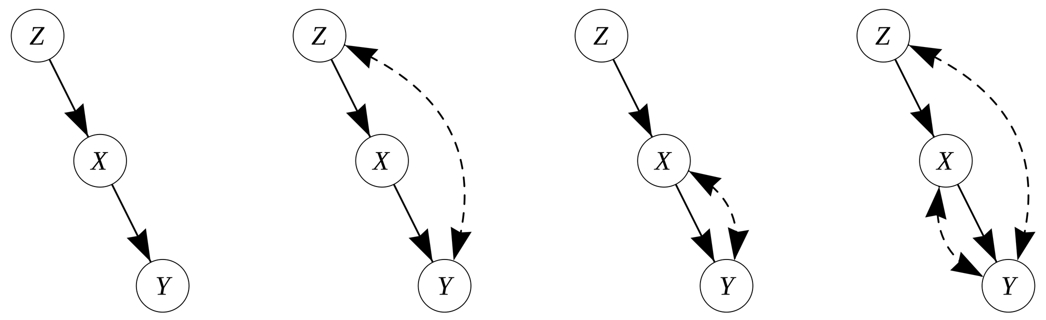
2. *Partial-orderedness* among Intervention Sets
A set of variables **X** may be preferred to another set of variables **Z** whenever their maximum achievable expected rewards can be ordered:
$$\mu_{\mathbf{x}^*} = \max_{\mathbf{x} \in D(\mathbf{X})} \mu_{\mathbf{x}} \geq \max_{\mathbf{z} \in D(\mathbf{Z})} \mu_{\mathbf{z}} = \mu_{\mathbf{z}^*}$$

— **Possibly-Optimal Minimal Intervention Set (POMIS, Def. 2)**

- Each MIS that can achieve an optimal expected reward in some SCM $\mathcal{M}$ confirming to the causal graph $\mathcal{G}$ is called a POMIS.
- Clearly, pulling non-POMISs will incur regrets and delay the identification of the optimal arms.

Toy Examples for MISs and POMISs

(* a dashed bidirected edge = existence of an unobserved confounder)



Same MISs $\{\emptyset, \{X\}, \{Z\}\}$ since $do(x)=do(x,z)$ for $z \in D(Z)$.
POMIS are $\{\{X\}\}$, $\{\emptyset, \{X\}\}$, $\{\{Z\}, \{X\}\}$, $\{\emptyset, \{Z\}, \{X\}\}$

- We characterized a complete condition whether an IS is a **(PO)MIS**.
- We devised an algorithmic procedure to enumerate all **(PO)MIS** given $\langle \mathcal{G}, Y \rangle$.

Empirical Evaluation

4 strategies $\times$ **2** base MAB solvers $\times$ **3** tasks; ($T = 10k$, 300 simulations)

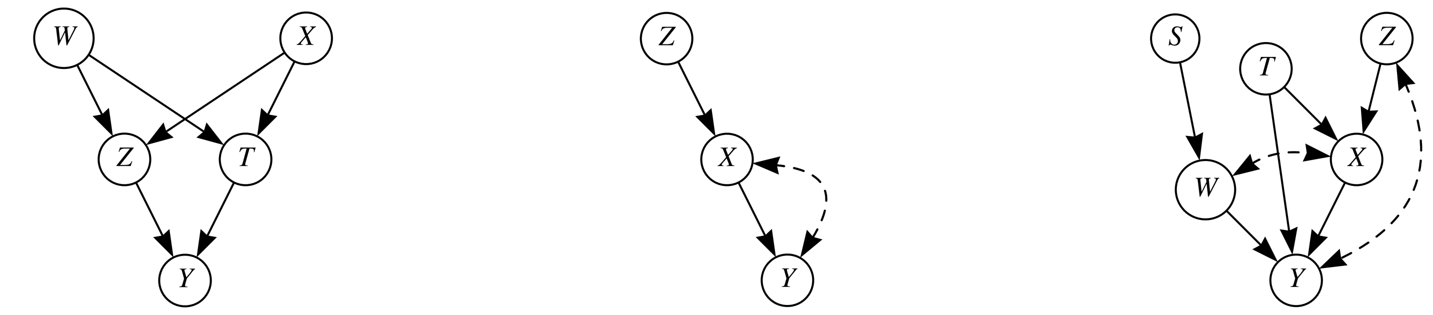
Strategies

- **Brute-force**: all possible arms, $\{\mathbf{x} \in D(\mathbf{X}) \mid \mathbf{X} \subseteq \mathbf{V} \setminus \{Y\}\}$ (aka all-subsets)
- **All-at-once**: intervene on all variables simultaneously, $D(\mathbf{V} \setminus \{Y\})$
- **MIS**: arms related to MISs
- **POMIS**: arms related to POMISs

Base MAB solvers

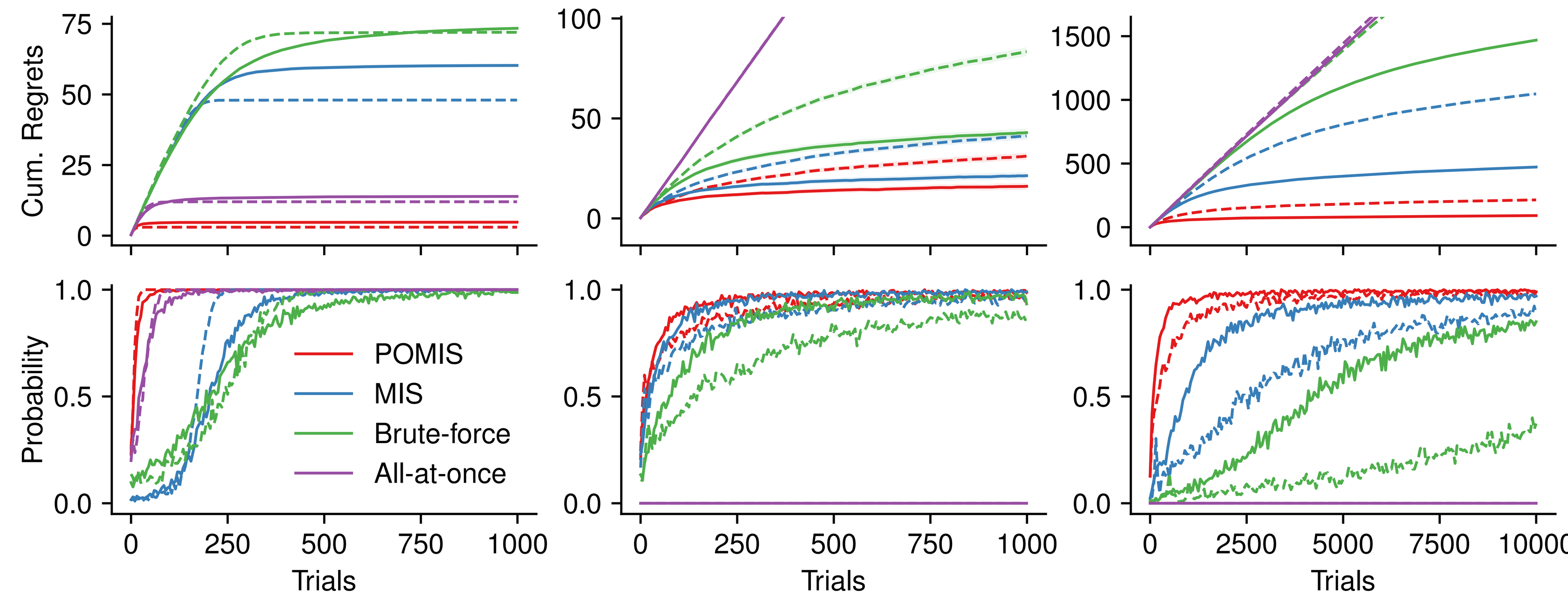
Thompson Sampling (TS) and kl-UCB

Tasks



Results

(**top**) averaged cumulative regrets and (**bottom**) optimal arm probability
TS in solid lines, kl-UCB in dashed lines



- CRs: **Brute-force** $\geq$ **MIS** $\geq$ **POMIS** (smaller the better)
- If the number of arms for **All-at-once** is *smaller* than **POMIS**, then, it implies that **All-at-once** is missing possibly-optimal arms.

Conclusions

- Introduced **SCM-MAB** = MAB + SCM = $\frac{\text{MAB}}{\text{SCM}}$.
- Characterized structural properties (equivalence, partial-orderedness) in SCM-MAB given a causal graph.
- Studied conditions under which intervening on a set of variables might be optimal (POMIS).
- Empirical results corroborate theoretical findings.

- We have a $\star$new$\star$ paper to be presented at **AAAI**'2019 !
  - Introduced **non-manipulability** constraints (not all variables are intervenable),
  - Characterized **MISs** / **POMISs** w/ the constraints,
  - Introduced novel strategy to leverage structural relationships across arms with improved finite-sample properties.

Papers at `causalai.net`
Code at `https://github.com/sanghack81/SCMMAB-NIPS2018`