Structural Causal Bandits under Markov Equivalence

Min Woo Park[1] Andi Ardity[2] Elias Bareinboim[2*] Sanghack Lee[1*] ([1]Seoul National University [2]Columbia University)

NEURAL INFORMATION PROCESSING SYSTEMS

Seoul National University Graduate School of Data Science

COLUMBIA UNIVERSITY IN THE CITY OF NEW YORK

Multi-Armed Bandits

Multi-armed bandit (MAB) problem is a classic sequential decision-making problem.

Arms a set of arms **A** to play; each arm associates with a reward distribution.

Play pulling an arm $A_{\mathbf{x}}$ for each round $t \in [T]$,

Reward a reward $Y_{\mathbf{x}}$ is drawn from the arm's reward distribution,

Goal to minimize a cumulative regret over a total round T.

Structural Causal Bandits - SCM-MAB

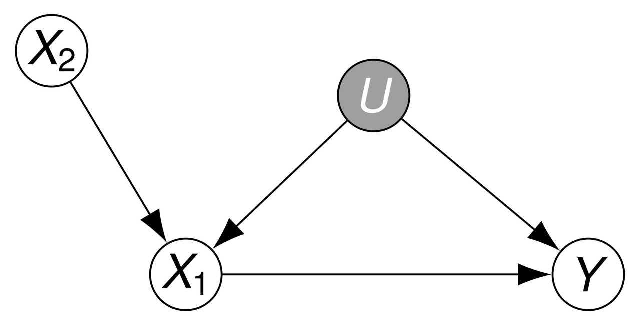
- Multi-armed bandit through Causal Lens.
- A Structural Causal Model (SCM) $\mathcal{M} = \langle \mathbf{U}, \mathbf{V}, \mathbf{F}, P(\mathbf{U}) \rangle$ (Pearl, 2000):
  - $\mathbf{U}$ unobserved variables;
  - $\mathbf{V}$ observed variables;
  - $\mathbf{F}$ causal mechanisms for $\mathbf{V}$ using $\mathbf{U}$ and $\mathbf{V}$;
  - $P(\mathbf{U})$ a joint distribution over $\mathbf{U}$ (**randomness**).
- SCM-MAB = MAB on SCM: a SCM $\mathcal{M}$; a reward variable $Y \in \mathbf{V}$.

  Intervention sets (ISs) correspond to *all* subset of $\mathbf{V} \setminus \{Y\}$.

  Arms **A** correspond to values for intervention sets. i.e., action space $\{A_{\mathbf{x}} \mid \mathbf{x} \in \mathcal{D}_{\mathbf{X}}, \mathbf{X} \subseteq \mathbf{V} \setminus \{Y\}\}$.

  Reward: distribution $P(Y_{\mathbf{x}}) := P(Y \mid do(\mathbf{x})) = P_{\mathbf{x}}(Y)$, expectation, $\mu_{\mathbf{x}} := \mathbb{E}[Y \mid do(x)]$.
- **Assumption**. a causal graph $\mathcal{G}$ of $\mathcal{M}$ is accessible.

Example. We can control 2 binary variables, X_1 and X_2.

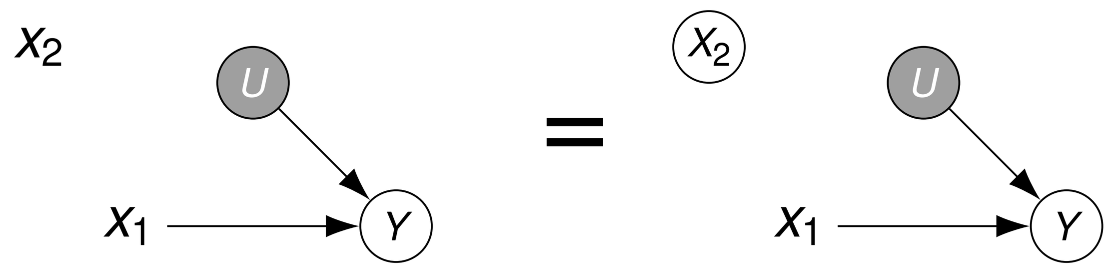


Intervention sets (4): $\emptyset$, $\{X_1\}$, $\{X_2\}$, $\{X_1, X_2\}$.

Arms (9): $\emptyset$, $do(x_1 = 0)$, $do(x_1 = 1)$, $\cdots$, $do(x_1 = 1, x_2 = 1)$.

Structural Properties of SCM-MAB

Property 1(Equivalence among Arms)**.** Two arms share the same reward distribution.



$$\mu_{x_1, x_2} = \mu_{x_1}$$

- **Minimal Intervention Set** (MIS): A minimal and ancestral set of variables among ISs sharing the same reward distribution.

Property 2(Partial-Orderedness)**.** Maximum achievable expected rewards can be ordered.

$$\mu_{\emptyset} = \sum_{x_2} \mu_{x_2} P(x_2) \leq \sum_{x_2} \mu_{x_2^*} P(x_2) = \mu_{x_2^*}.$$

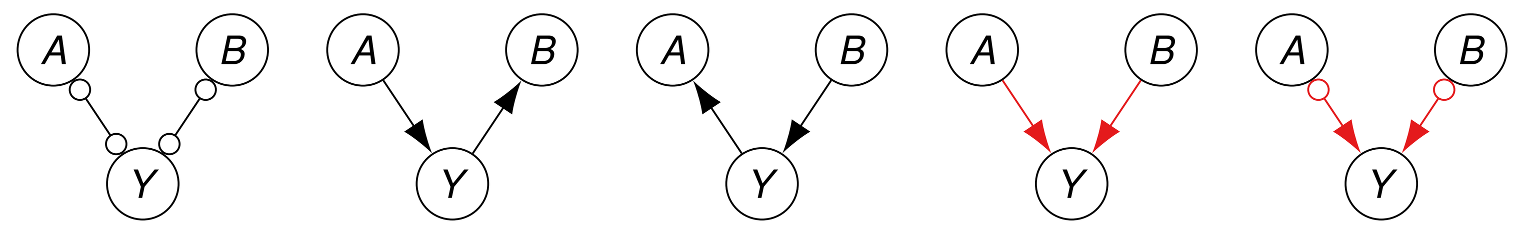
- **Possibly-Optimal Minimal Intervention Set** (**POMIS**): An MIS that can achieve an optimal expected reward in some SCM $\mathcal{M}$ conforming to the causal graph $\mathcal{G}$ is called a POMIS.

Inclusion: All ISs $\supseteq$ MIS $\supseteq$ POMIS.

Cumulative Regrets: All ISs $\geq$ MIS $\geq$ POMIS (smaller the better).

Markov Equivalence Class

- SCM-MAB assumes an access to the causal graph $\mathcal{G}$ of $\mathcal{M}$.
- One can discover only a *partial ancestral graphs* (PAG) which represents Markov equivalence classes (MEC).



(a) $\mathcal{P}$ (b) $\mathcal{G}_1$ (c) $\mathcal{G}_2$ (d) $\mathcal{G}_3$ (e) $\mathcal{P}'$

where the causal graphs $\mathcal{G}_1$ and $\mathcal{G}_2$ are represented by $\mathcal{P}$ and they share the conditional independence $A \perp\!\!\!\perp B \mid Y$. In contrast, $\mathcal{G}_3$ is not represented by the PAG $\mathcal{P}$ but rather by $\mathcal{P}'$.

- We extend the SCM-MAB in the context of PAGs.

SCM-MAB under MEC

- **Assumption**. a PAG $\mathcal{P}$ of $\mathcal{M}$ is accessible.
- **Definitely MIS** (**DMIS**): A minimal and definitely ancestral set of variables among ISs sharing the same reward distribution.

  DMIS is similar to MIS, but it considers *undetermined ancestral relations*, e.g., A and B cannot both be ancestors of Y simultaneously, since such causal diagrams correspond to another PAG (MEC), i.e., $A \not\perp\!\!\!\perp B \mid Y$.

  $\{A, B\}$ is *not* a DMIS given $\mathcal{P}$.
- **Possibly-Optimal Minimal Intervention Set** (**POMIS**): An **DMIS** that can achieve an optimal expected reward in some SCM $\mathcal{M}$ conforming to the PAG $\mathcal{P}$.

  Inclusion: All ISs $\supseteq$ DMIS $\supseteq$ POMIS.

  Cumulative Regrets: All ISs $\geq$ DMIS $\geq$ POMIS.

Identifying POMIS Algorithm

Input: PAG $\mathcal{P}$; Reward variable Y; Intervention set $\mathbf{X}$.

Output: whether $\mathbf{X}$ is a POMIS.

1. Check whether the given $\mathbf{X}$ is a DMIS.
2. If $\mathbf{X}$ is a DMIS, then certain edge orientations can be determined. Incorporate this information into the PAG.
3. Enumerate all possible orientation combinations around $\mathbf{X} \cup \{Y\}$ and test POMIS conditions.

Conclusion

1. Given a PAG, you do not need to enumerate all causal diagrams conforming to the PAG to compute POMIS!
2. One can simply use only the arms corresponding to POMISs with any standard solver (TS, UCB).

Acknowledgements: This work was supported in part by the NSF, ONR, AFOR, DOE, Amazon, JP Morgan, and The Alfred P. Sloan Foundation.
Min Woo Park and Sanghack Lee were partly supported by the IITP (RS-2022-II220953/25%, RS-2025-02263754/25%) and
NRF (RS-2023-00211904/25%, RS-2023-00222663/25%) grant funded by the Korean government.