Structural Causal Bandits with Non-manipulable Variables

Sanghack Lee and Elias Bareinboim (CausalAI Lab, Purdue University)

Multi-Armed Bandit

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 $A_{\mathbf{x}} \in \mathbf{A}$ for each round,
- **Reward**: a reward $Y_{\mathbf{x}}$ is drawn from the arm's reward distribution,
- **Goal**: to minimize a cumulative regret over T.

Structural Causal Model — *the* Causal Framework

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

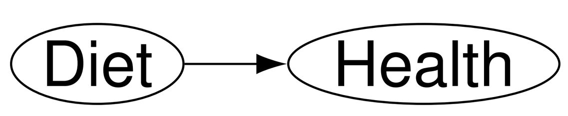
- **U**: **unobserved** variables;
- **V**: **observed** variables;
- **F**: **causal mechanisms** for **V** using **U** and **V**;
- $P(\mathbf{U})$: a joint distribution over **U** (**randomness**).

SCM-MAB = MAB on SCM

- $\langle \mathcal{M}, Y, \mathbf{N} \rangle$: a SCM $\mathcal{M}$; a reward variable $Y \in \mathbf{V}$; non-manipulables $\mathbf{N}$
- Arms **A** correspond to *all* possible interventions $\{A_{\mathbf{x}} \mid \mathbf{x} \in D(\mathbf{X}), \mathbf{X} \subseteq \mathbf{V} \setminus \mathbf{N} \setminus \{Y\}\}$.
- Reward: distribution $P(Y_{\mathbf{x}}) := P(Y \mid do(\mathbf{x})) = P_{\mathbf{x}}(Y)$, expectation, $\mu_{\mathbf{x}} := \mathbb{E}[Y \mid do(\mathbf{x})]$.

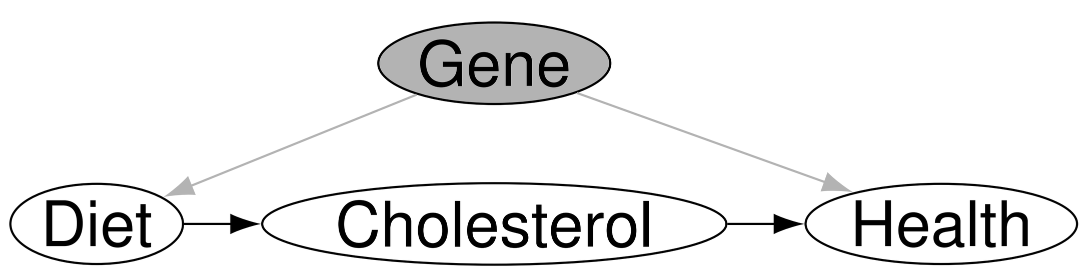
Assumption: 1) a causal graph $\mathcal{G}$ of $\mathcal{M}$ is accessible; 2) values of observable variables, **v**, are obtained for each play.

MAB



Arms = a set of diet values

SCM-MAB



Arms = doing nothing, values for diet, cholesterol, and both (combinations).
With $\mathbf{N} = \{Cholesterol\}$, Arms = doing nothing, values for diet

*valid question: Can't we just use MAB with D,C → Health and 'do-nothing' arm?

Structural Properties of SCM-MAB — How can we utilize the given causal structure? dependency among the arms?

1. Equivalence

Two arms share the same reward distribution, e.g.,

$$\mu_{d,c} = \mu_c$$

whenever intervening on some variables doesn't have a causal effect on the outcome.

→ Test $P(y \mid do(d, c)) = P(y \mid do(c))$ through $Y \perp\!\!\!\perp C \mid D$ in $\mathcal{G}_{\overline{\{D,C\}}}$ (Rule 3 of *do*-calculus, Pearl (2000)).

Minimal Intervention Set (MIS)

- 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

A set of variables **X** may be preferred to another set of variables **Z** whenever their maximum achievable expected rewards can be ordered:

$$\mu_{c^*} = \max_c \mu_c \geq \max_d \mu_d = \mu_{d^*}$$

$$\mu_d = \sum_c \mu_c P(c|d) \leq \sum_c \mu_{c^*} P(c|d) = \mu_{c^*}$$

Possibly-Optimal MIS (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.
- Clearly, pulling non-POMISs will incur regrets and delay the identification of the optimal arms.

3. Identifiability

Can one arm's reward distribution $P_{\mathbf{x}}(y)$ be expressed with other arms' distributions?

$$P_d(y) = \sum_c P(c|d) \sum_{d'} P(y|c, d') P(d')$$

z²ID algorithm:
outputs an expression (if it can) given a query (i.e., reward distribution) and available distributions.

Minimum Variance Weighting:
is a principled way to combine estimates from multiple estimators using multiple data sources

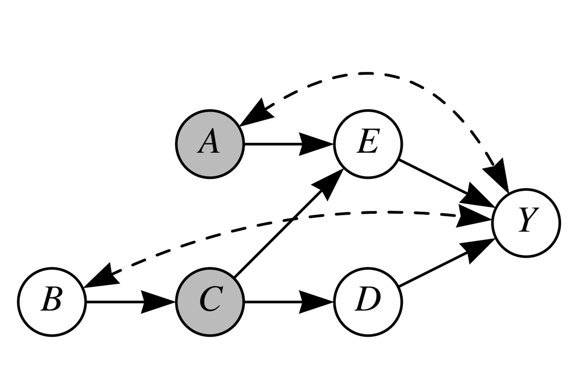
SCM-MAB algorithms

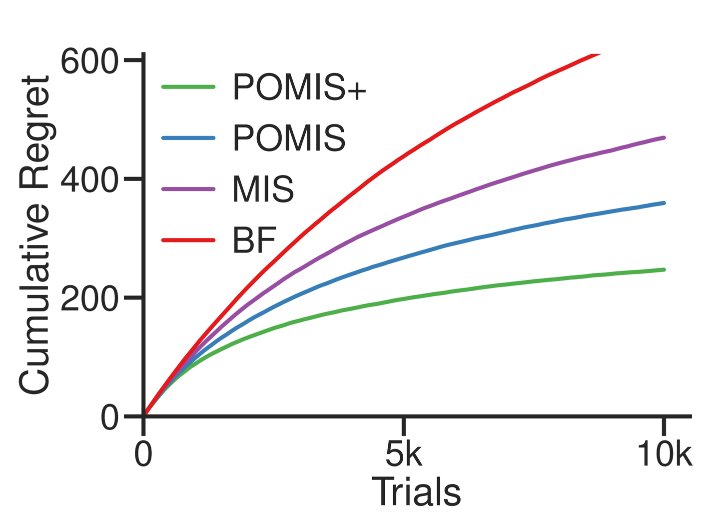
1. Play only POMIS arms (→ small # of arms)
2. Incorporate **z²ID** and **MVW** (→ more accurate estimation)

Empirical Evaluation

- **4** strategies: Brute-force (all ISs), MIS, POMIS, POMIS+
- **2** base MAB algorithms: TS, kl-UCB
- **3** SCM-MAB problems (w/ binary variables)

e.g.,





Performance: POMIS+ > POMIS ≥ MIS ≥ Brute-force
⋆ Note that POMISs ⊆ MISs ⊆ all ISs

Conclusions

- Causal mechanisms do exist.
- Agents ignorant to an underlying causal mechanism might behave suboptimally.

defined: **SCM-MAB** w/ non-manipulability constraints
studied: **3 structural properties** of SCM-MAB
devised: SCM-MAB **algos** w/ the structural properties
observed: better performance than MAB algo w/o causal knowledge

Visit `causalai.net` for more papers on causality.