Stochastic Intervention with Large Action Space

BIS631 Advance Topics in Causal Inference

Tatsuhiro Shimizu

Agenda

1. Introduction

2. Setting of the Problem
   - Data
   - Problem

3. Literature Review
   - Direct Method (DM)
   - Inverse Propensity Score (IPS)
   - Doubly Robust (DR)
   - Marginalized Inverse Propensity Score (MIPS)

4. New Estimator: Marginalized Doubly Robust (MDR)
   - Definition of MDR
   - Theoretical Guarantees of MDR
   - Simulation study

5. Limitation and Future Work

Introduction

Introduction

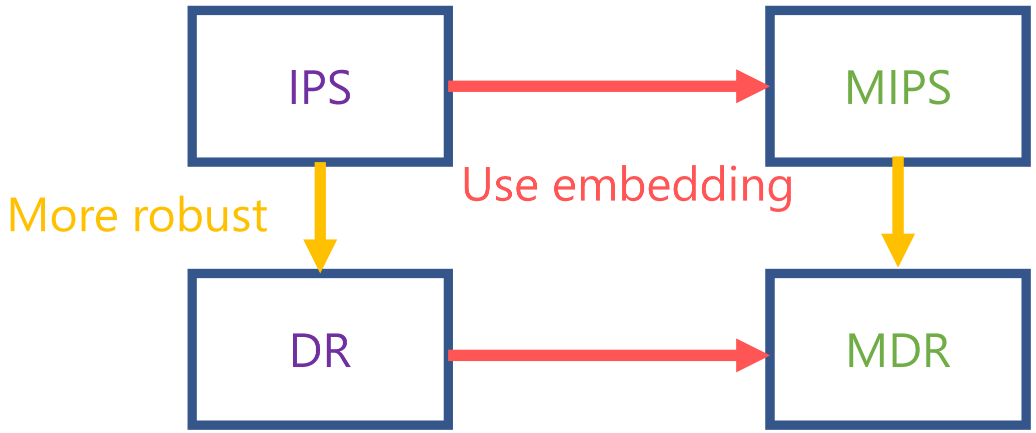
- We generalize the theory in Causal Inference with stochastic intervention

- Why stochastic intervention?
  1. For risk aversion
     - Variance of the effect of deterministic intervention is different
       - Example in Finance: Even if the expected return is the highest stock, we try to have the optimal portfolio by having multiple different stocks
  2. For exploration in Online Learning

- We can use the classical Causal Inference methods
  - ➢E.g., Inverse Propensity Score (IPS) and Doubly Robust (DR)

Introduction

- We consider the irregular setting where we have lots of deterministic interventions
  - ➢IPS and DR are not good estimators due to the high variance
  - ➢Marginalized Inverse Propensity Score (MIPS) works well using the embedding of the action
  - ➢I introduce Marginalized Doubly Robust (MDR) to achieve better estimator



Setting of the Problem

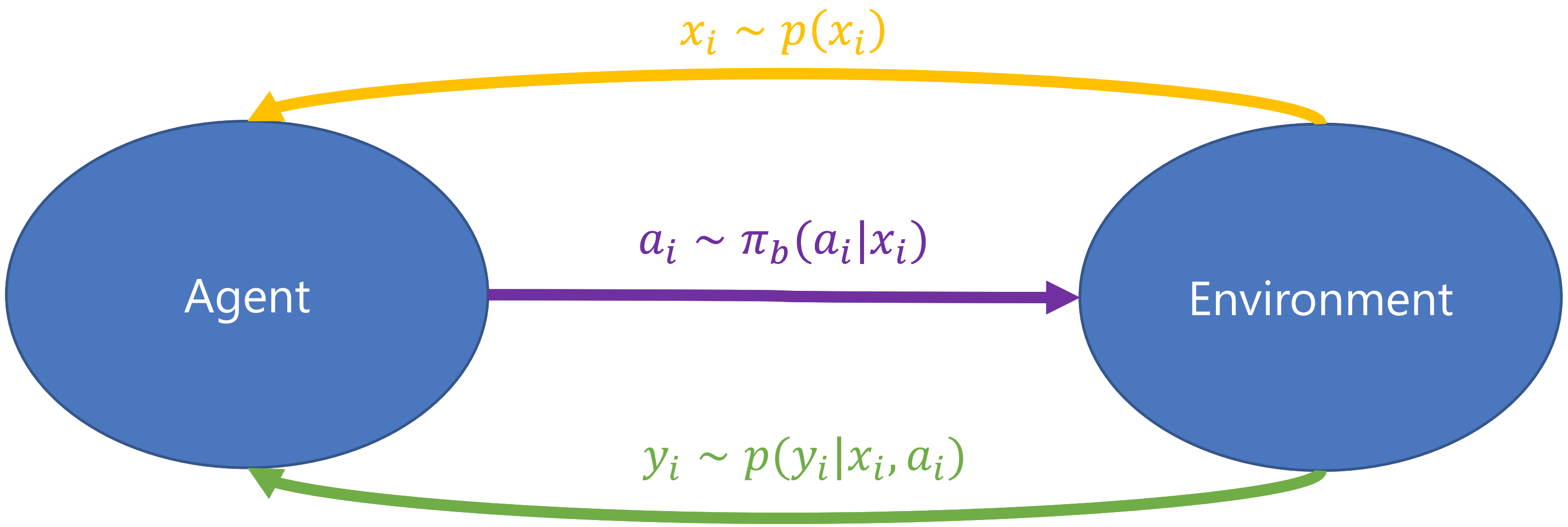
Setting of Problem: Data

- We consider the data
  - Contextual vector (covariate): $x \in \mathcal{X} = \mathbb{R}^{d_x}$
  - Action (deterministic intervention): $a \in \mathcal{A}$
  - Outcome: $y \in [0, y_{max}]$

- Data generating Process
  - $x \sim p(x)$ where $p(x)$ is an unknown distribution
  - $a \sim \pi(a|x)$ where $\pi: \mathcal{X} \rightarrow \Delta(\mathcal{A})$ is the stochastic intervention called policy
  - $y \sim p(y|x, a)$ where $p(y|x, a)$ is an unknown distribution

Setting of Problem: Data

- Observed Data
  - n units
  - $\mathcal{D} = \{(x_i, a_i, y_i)\}_{i \in [n]}$
    - For each unit $i \in [n]$, we observe i.i.d.
      - $x_i \sim p(x_i)$
      - $a_i \sim \pi_b(a_i|x_i)$ where π_b is the already used policy in the system called behavior policy
      - $y_i \sim p(y_i|x_i, a_i)$



Setting of Problem: Problem

- We define how good a policy π is
  - Definition of Value function $V(\pi)$
    - $V(\pi) \coloneqq \mathbb{E}_{p(x)\pi(a|x)p(y|x,a)}[y] = \mathbb{E}_{p(x)\pi(a|x)}[q(x,a)]$
    - Where $q(x,a) = \mathbb{E}_{p(y|x,a)}[y|x,a]$ is the expected outcome given x and a

- If we know $p(x)$ and $p(y|x,a)$, then we can find the best policy

- As we do not know $p(x)$ and $p(y|x,a)$, we construct estimator
$$\hat{V}(\pi_e;\mathcal{D}) \approx V(\pi_e)$$

- Use Mean Squared Error (MSE) as the metric of how good the estimator is
$$\mathrm{MSE}\left(\hat{V}(\pi_e;\mathcal{D})\right) = \mathbb{E}_{\mathcal{D}}\left[\left(V(\pi) - \hat{V}(\pi_e;\mathcal{D})\right)^2\right]$$

Literature Review

Literature Review: Direct Method (DM)

- Direct Method (DM) (Beygelzimer and Langford 2009)

$$\hat{V}_{\mathrm{DM}}(\pi_e; \mathcal{D}, \hat{q}) := \frac{1}{n} \sum_{i \in [n]} \mathbb{E}_{\pi_e(a|x_i)} [\hat{q}(x_i, a)]$$

  - where $\hat{q}(x, a)$ is the estimated expected outcome

$$\hat{q} \leftarrow \underset{q' \in \mathcal{Q}}{\operatorname{argmin}} \left\{ \frac{1}{n} \sum_{i \in [n]} \left(y_i - q'(x_i, a_i) \right)^2 \right\}$$

  - Unbiased under the unbiasedness of $\hat{q}$

Literature Review: Inverse Propensity Score (IPS)

- Inverse Propensity Score (IPS) (Horvitz and Thompson 1952)

$$\hat{V}_{\mathrm{IPS}}(\pi_e; \mathcal{D}) \coloneqq \frac{1}{n} \sum_{i \in [n]} w(x_i, a_i) y_i$$

  - Where the importance weight $w(x_i, a_i)$ is

$$w(x_i, a_i) \coloneqq \frac{\pi_e(a_i|x_i)}{\pi_b(a_i|x_i)}$$

  - Unbiased under the common support
    - Common Support: $\pi_e(a|x) > 0 \implies \pi_b(a|x) \ \forall x \in \mathcal{X}, a \in \mathcal{A}$
  - Variance of IPS (Saito and Joachims 2022) is

$$n\mathbb{V}_{\mathcal{D}}\left[\hat{V}_{\mathrm{IPS}}(\pi_e; \mathcal{D})\right] = \underbrace{\mathbb{E}_{p(x)\pi_b(a|x)}\left[w(x,a)^2 \sigma(x,a)^2\right]}_{\text{large when } |\mathcal{A}| \gg 1} + \mathbb{V}_{p(x)}\left[\mathbb{E}_{\pi_b(a|x)}[w(x,a)q(x,a)]\right] + \underbrace{\mathbb{E}_{p(x)}\left[\mathbb{V}_{\pi_b(a|x)}[w(x,a)q(x,a)]\right]}_{\text{large when } |\mathcal{A}| \gg 1}$$

Literature Review: Doubly Robust(DR)

- Doubly Robust(DR) (Dudik et al 2011, Cassel et al 1976)

$$\hat{V}_{\mathrm{DR}}(\pi_e; \mathcal{D}; \hat{q}) \coloneqq \frac{1}{n}\sum_{i\in[n]}\{\mathbb{E}_{\pi_e(a|x_i)}[\hat{q}(x_i, a)] + w(x_i, a_i)\big(y_i - \hat{q}(x_i, a_i)\big)\}$$

  - ➢Unbiased under the common support or unbiasedness of $\hat{q}$
  - ➢Variance of DR (Huang et al 2021) is

$$n\mathbb{V}_{\mathcal{D}}\big[\hat{V}_{\mathrm{DR}}(\pi_e; \mathcal{D}, \hat{q})\big] =$$

$$\underbrace{\mathbb{E}_{p(x)\pi_b(a|x)}[w(x,a)^2\sigma(x,a)^2]}_{\text{large when } |\mathcal{A}|\gg 1} + \mathbb{V}_{p(x)}\left[\mathbb{E}_{\pi_b(a|x)}[w(x,a)q(x,a)]\right] + \underbrace{\mathbb{E}_{p(x)}\left[\mathbb{V}_{\pi_b(a|x)}[w(x,a)\Delta(x,a)]\right]}_{\text{large when } |\mathcal{A}|\gg 1}$$

where$\Delta(x,a) = q(x,a) - \hat{q}(x,a)$ is the error of the estimation of expected outcome

Literature Review: Marginalized Inverse Propensity Score (MIPS)

- To overcome the high variance of IPS and DR
- MIPS use the embedding $e \in \mathcal{E} \subset \mathbb{R}^{d_e}$ of the action $a \in \mathcal{A}$ for the importance weight

- Example
  - Want to construct the optimal movie recommendation system (e.g. Netflix)
    - Action a: movies
    - Action embedding e: movie genres, actors, director

- To use MIPS, we need the modified data generating process
  - Context vector (covariates) $x \sim p(x)$
  - Action $a \sim \pi(a|x)$
  - Action embedding $e \sim p(e|x, a)$
  - Outcome $y \sim p(y|x, a, e)$

Literature Review: Marginalized Inverse Propensity Score (MIPS)

- New observed data
  - $\mathcal{D} = \{(x_i, a_i, e_i, y_i)\}_{i \in [n]}$
    - For each unit $i \in [n]$, we observe i.i.d.
      - $x_i \sim p(x_i)$
      - $a_i \sim \pi_b(a_i|x_i)$
      - $e_i \sim p(e_i|x_i, a_i)$
      - $y_i \sim p(y_i|x_i, a_i, e_i)$
- New value function $V(\pi)$

$$\begin{aligned} V(\pi) &:= \mathbb{E}_{p(x)\pi(a|x)p(e|x,a)p(y|x,a,e)}[y] \\ &= \mathbb{E}_{p(x)\pi(a|x)p(e|x,a)}[q(x, a, e)] \\ &= \mathbb{E}_{p(x)\pi(a|x)}[q(x, a)] \end{aligned}$$

Literature Review: Marginalized Inverse Propensity Score (MIPS)

- Marginalized Inverse Propensity Score (MIPS) (Saito and Joachims 2022)

$$\hat{V}_{\mathrm{MIPS}}(\pi_e; \mathcal{D}) := \frac{1}{n} \sum_{i \in [n]} w(x_i, e_i) y_i$$

  - ➢ Where the marginalized importance weight $w(x_i, e_i)$ is

$$w(x_i, e_i) := \frac{p(e|x, \pi_e)}{p(e|x, \pi_b)}$$

  - ➢ Where $p(e|x, \pi) = \sum_{a \in \mathcal{A}} \pi(a|x) p(e|x, a)$ is the marginal distribution of embedding

  - ➢ Unbiased under 1. no direct effect and 2. common embedding support
    - No direct effect: a and y is independent given x and e
    - Common embedding support: $p(e|x, \pi_e) > 0 \implies p(e|x, \pi_b) > 0 \quad \forall x \in \mathcal{X}, e \in \mathcal{E}$

- If we can find the good representation of action, then we can have the lower variance than IPS and DR

New Estimator: Marginalized Doubly Robust (MDR)

New Estimator: Marginalized Doubly Robust (MDR)

- Idea: Combine MIPS and DR to obtain the better estimator of $V(\pi)$
- Marginalized Doubly Robust (MDR)

$$\hat{V}_{\mathrm{MDR}}(\pi_e; \mathcal{D}; \hat{q}) := \frac{1}{n} \sum_{i \in [n]} \{\mathbb{E}_{\pi_e(a|x_i)}[\hat{q}(x_i, a)] + w(x_i, e_i)\big(y_i - \hat{q}(x_i, a_i, e_i)\big)\}$$

- ➢Unbiased under
  1. the no direct effect
  2. either common embedding support or unbiasedness of $\hat{q}(x, a, e)$
- ➢More robust than MIPS

New Estimator: Marginalized Doubly Robust (MDR)

- Variance reduction of MDR against DR

$$
\begin{aligned}
& n\mathbb{V}_{\mathcal{D}}\left[\hat{V}_{\mathrm{DR}}(\pi_e;\mathcal{D},\hat{q})\right] - n\mathbb{V}_{\mathcal{D}}\left[\hat{V}_{\mathrm{MDR}}(\pi_e;\mathcal{D},\hat{q})\right] \\
&= \mathbb{E}_{\bar{d}_{\pi_b}}\left[w(x,a)^2\Delta(x,a)^2 - w(x,e)^2\Delta(x,a,e)^2\right]
\end{aligned}
$$

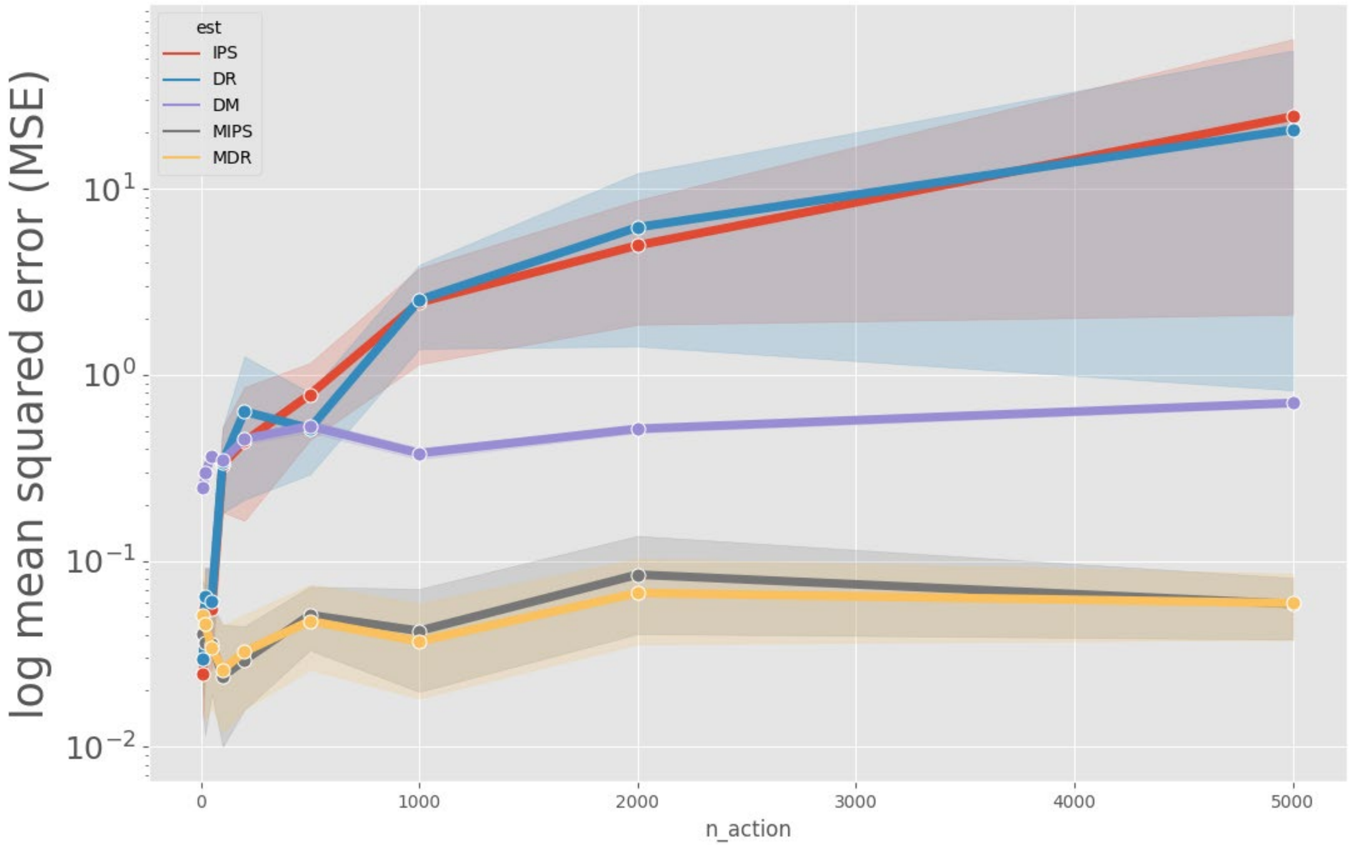
  - where
    - $\bar{d}_{\pi_b} \coloneqq p(x)\pi_b(a|x)p(e|x,a)$ is the visitation measure
    - $\Delta(\mathrm{x},\mathrm{a}) = \mathrm{q}(\mathrm{x},\mathrm{a}) - \hat{\mathrm{q}}(x,a)$ is the estimation error of expected outcome given x and a
    - $\Delta(\mathrm{x},\mathrm{a},\mathrm{e}) = \mathrm{q}(\mathrm{x},\mathrm{a},\mathrm{e}) - \hat{\mathrm{q}}(x,a,e)$ is the estimation error of expected outcome given x, a, and e

- If the embedding e represents action a well, then $w(x,a) > w(x,e)$

$$
n\mathbb{V}_{\mathcal{D}}\left[\hat{V}_{\mathrm{DR}}(\pi_e;\mathcal{D},\hat{q})\right] > n\mathbb{V}_{\mathcal{D}}\left[\hat{V}_{\mathrm{MDR}}(\pi_e;\mathcal{D},\hat{q})\right]
$$

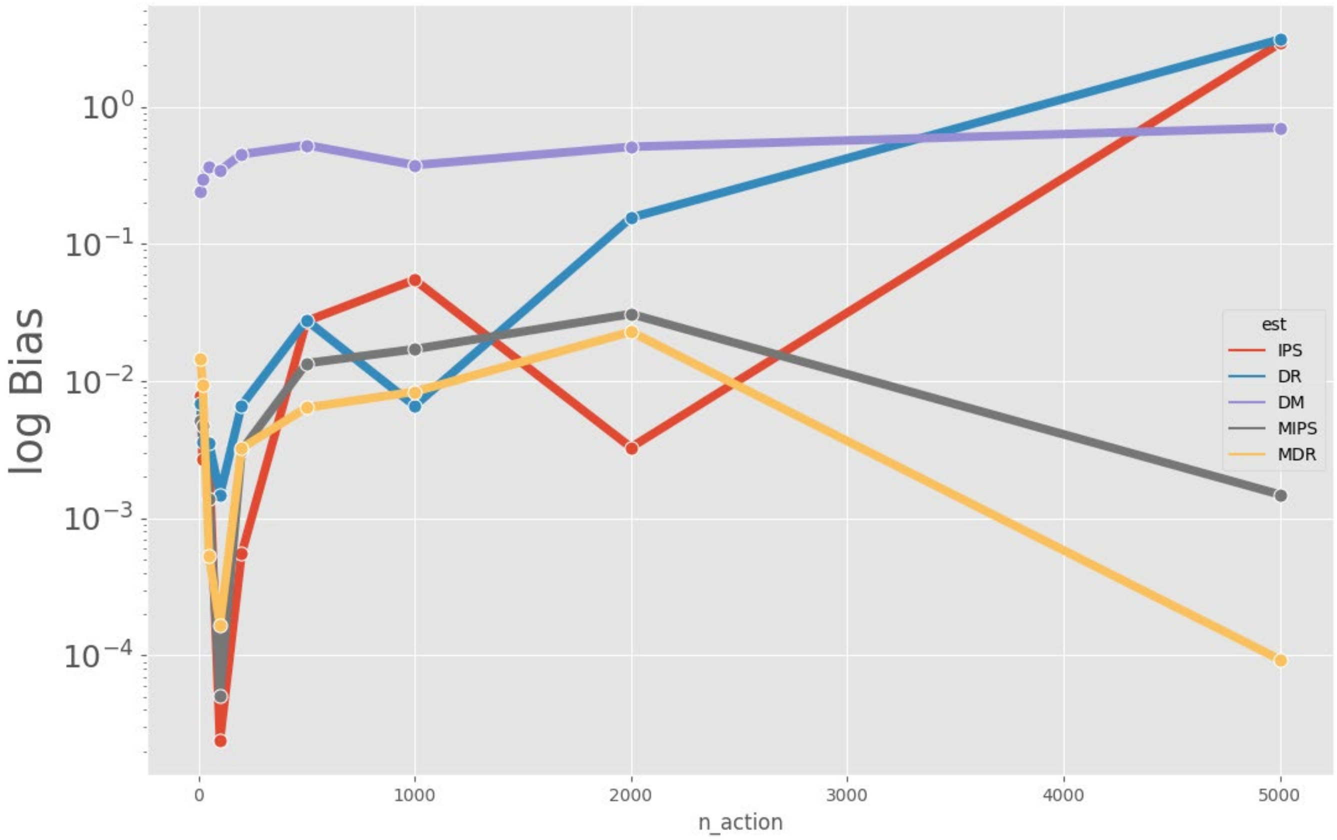
New Estimator: Marginalized Doubly Robust (MDR)

- Simulation study (MSE)
  - ➤MDR is the best of all estimators DM, IPS, DR, and MIPS



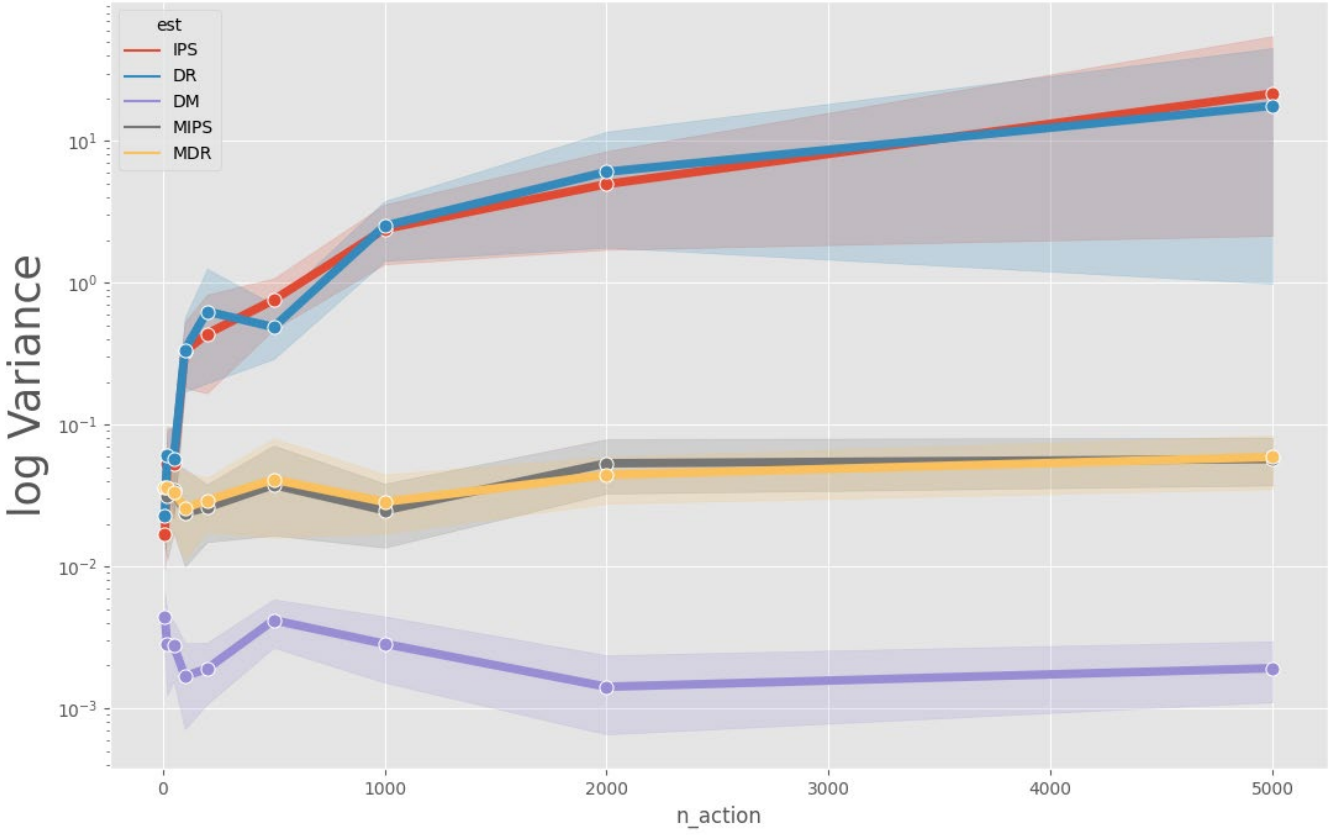
New Estimator: Marginalized Doubly Robust (MDR)

- Simulation study (Bias)
  - ➢MDR improve the bias of MIPS by the doubly robust properties



New Estimator: Marginalized Doubly Robust (MDR)

- Simulation study (Variance)
  - ➢MDR has the variance reduction against DR



Limitation and Future Work

Limitation and Future Work

- Limitation
  - ➢Simply used the exactly same setting for the simulation study(Saito and Joachism 2022)
  - ➢Did not cover how to find the better embedding e

- Future Work
  - ➢Empirically analyze how robust MDR is against the violation of the assumptions
  - ➢Construct the comprehensive algorithm or way to find the best embedding of action

References

- A. Beygelzimer and J. Langford. The offset tree for learning with partial labels. KDD, pp. 129–138, 2009. doi: https://doi.org/10.48550/arXiv.0812.4044.
- D. G. Horvitz and D. J. Thompson. A generalization of sampling without replacement from a finite universe. Journal of the American Statistical Association, 47(260):663–685, 1952. doi: http://www.jstor.org/stable/2280784.
- Miroslav Dudik, John Langford, and Lihong Li. Doubly robust policy evaluation and learning. ICML 2011, arXiv:1103.4601 [cs.LG], 2011. doi: https://doi.org/10.48550/arXiv.1103.4601.
- Yuta Saito and Thorsten Joachims. Off-policy evaluation for large action spaces via embeddings. ICML 2022, arXiv:2202.06317 [cs.LG], 2022. doi: https://doi.org/10.48550/arXiv.2202.06317.
- A. Strehl, J. Langford, L. Li, and S. Kakade. Learning from logged implicit exploration data. NuerIPS, pp. 2217–2225, 2011. doi: https://hunch.net/~jl/projects/interactive/scavenging/scavenging.pdf.
- J. Langford, A. L. Strehl, and J. Wortman. Exploration scavenging. ICML, pp. 528–535, 2008. doi: https://hunch.net/~jl/projects/interactive/scavenging/scavenging.pdf.
- C. M. Cassel, C. E. S¨arndal, and J. H. Wretman. Some results on generalized difference estimation and generalized regression estimation for finite populations. Biometrika, 63:615–620, 1976.
- A. Huang, L. Leqi, Z. C. Lipton, and K. Azizzadenesheli. Off-policy risk assessment in contextual bandits. NeurIPS 2021, arXiv:2104.08977 [cs.LG], 2021. doi: https://doi.org/10.48550/arXiv.2104.08977.
- Y. Saito, S. Aihara, M. Matsutani, and Y. Narita. Open bandit dataset and pipeline: Towards realistic and reproducible off-policy evaluation. arXiv, arXiv:2008.07146 [cs.LG], 2020. doi: https://doi.org/10.48550/arXiv.2008.07146.10