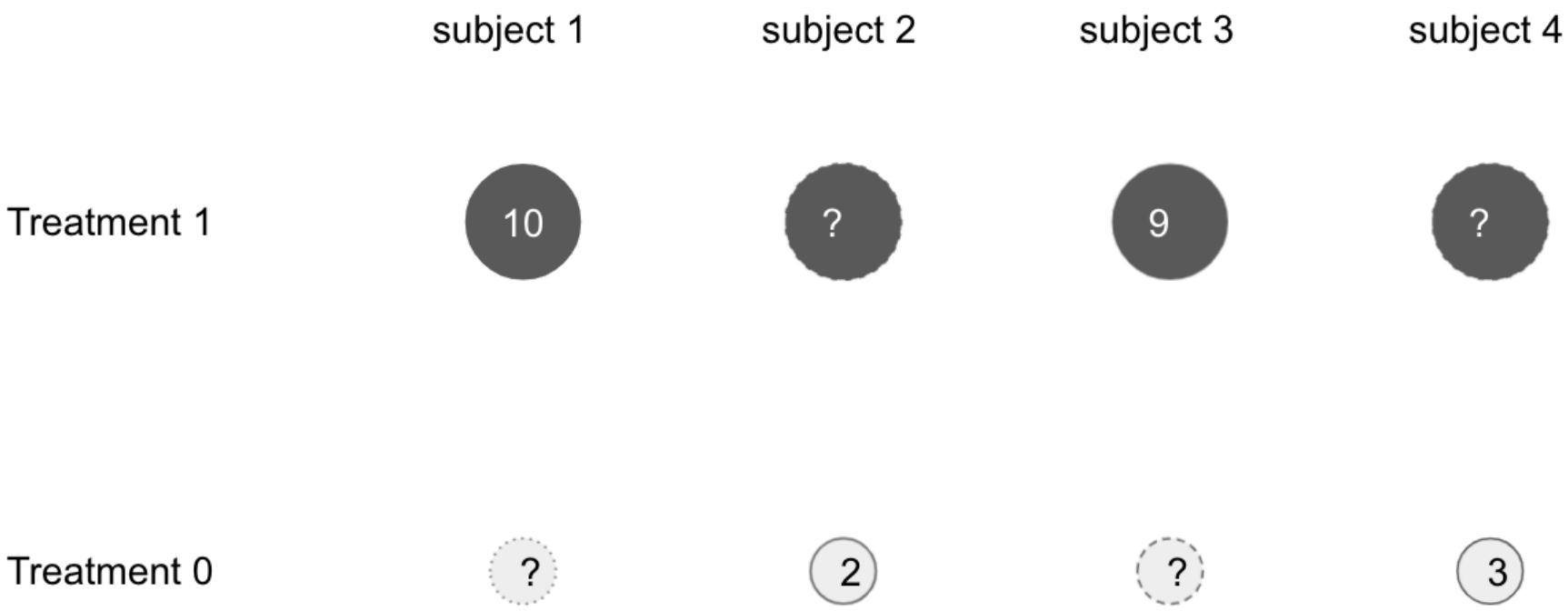
Kosorok Lab Journal Club
Kallus (2021). "More Efficient Policy Learning via Optimal Retargeting." *JASA*

Hunyong Cho

September 2021

A motivation of retargeting

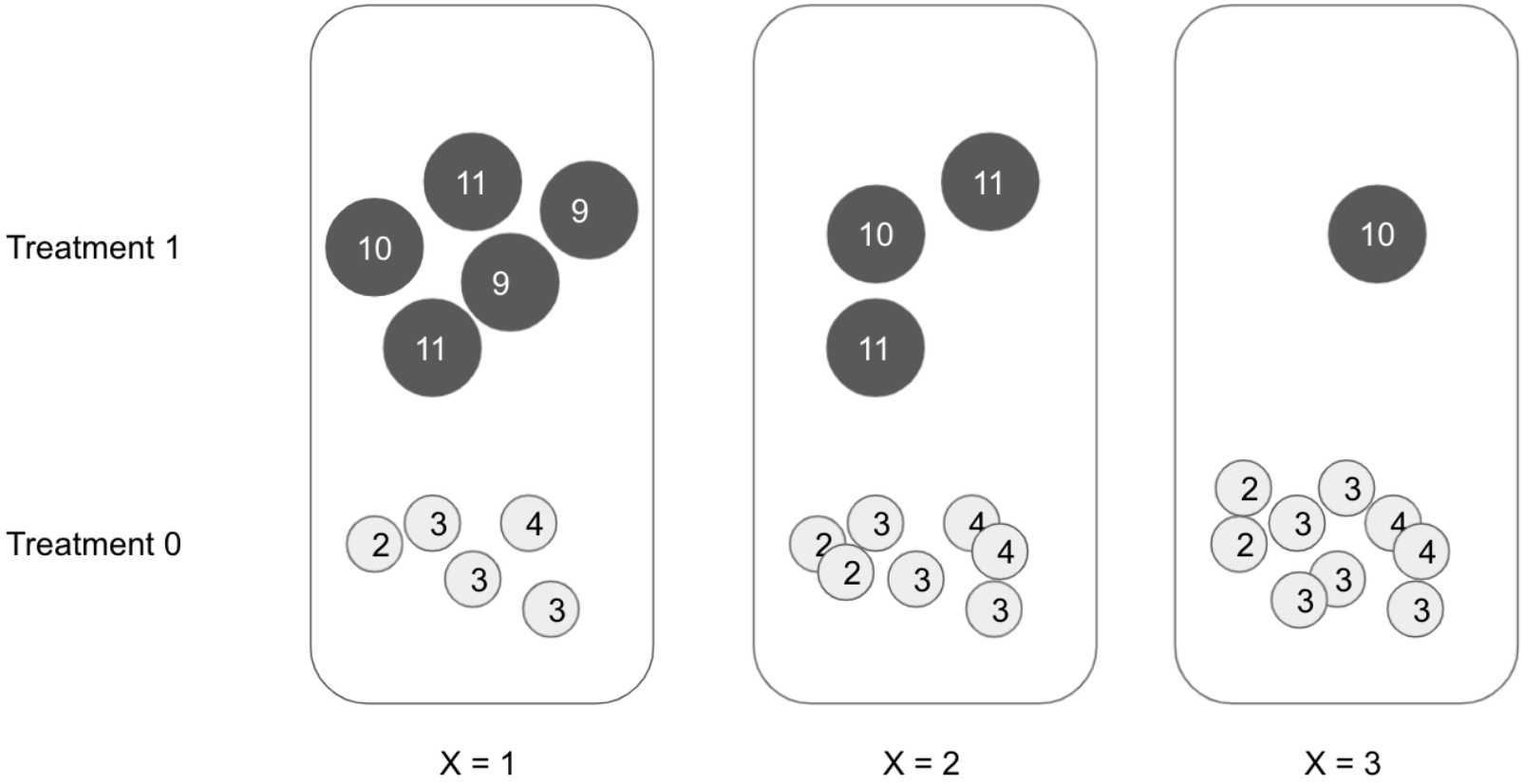
Counterfactual means and inverse probability weighting



$\hat{\tau} = \hat{\mu}_1 - \hat{\mu}_2 = 10 - 3$. $\hat{\tau} = \hat{\mu}_1 - \hat{\mu}_2 = 9.5 - 2.5$.

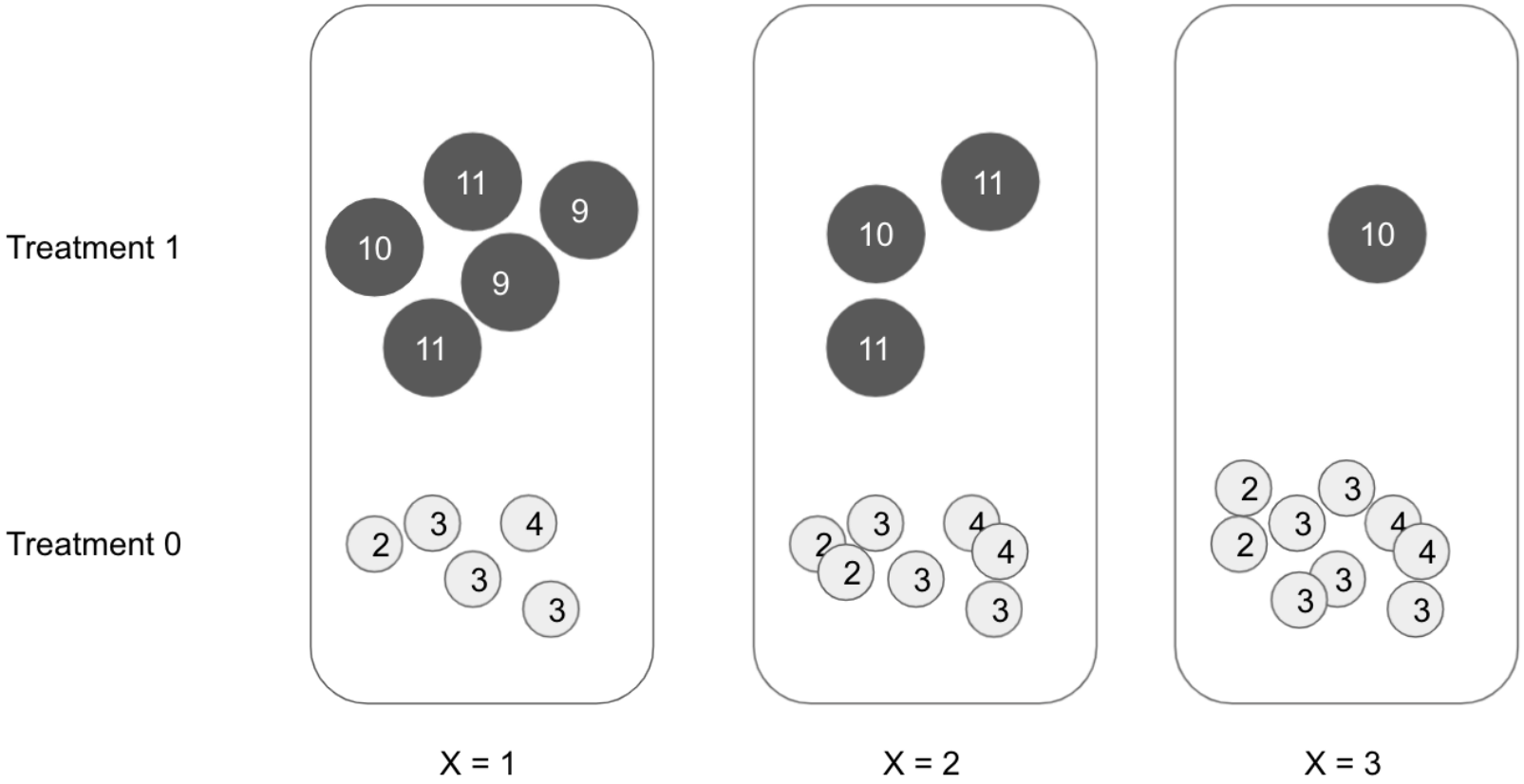
Each observed value represents two to cover the complete data.

Counterfactual conditional (on X) means



- Weights for $X = 3$ are $10/1$ for Tx 1 and $10/9$ for Tx 0.
- $\hat{\mu}^1(X = 3) = 10$ relies on a *single* observation. Instability!
- Sheds light on the problem of *non-overlap*.

Retargeting

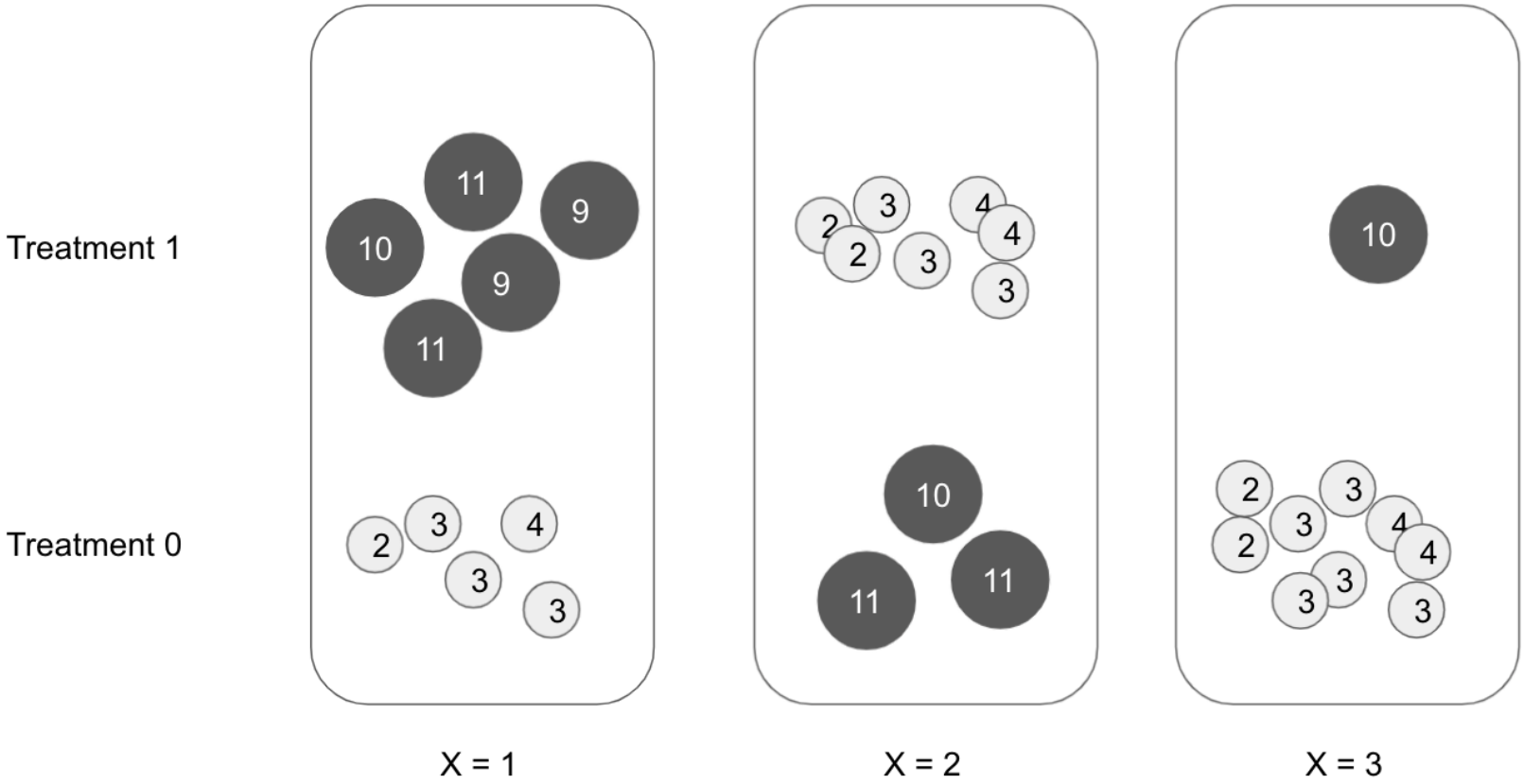


- Since Group $X = 3$ is less reliable, we want to give it less weights. E.g., $\phi(X)\{1 - \phi(X)\}$

"Retargeting"

- Does this retargeting affect optimization?

Retargeting



- For unrestricted Π, $\pi(x) = 1(\mu^1(x) > \mu^0(x))$ optimizes the retargeted value regardless of the weight.
- Let the restricted Π_0 be a set of monotone classifiers $\{1(x \geq b), 1(x < b) : b = 1, 2, 3, 4\}$,

Retargeted value

$$R(\pi; w, \rho) := \mathbb{E}\Big[w(X) \sum_{a\in\{0,1\}} \Big\{\pi(a|X) - \rho(a|X)\Big\}\mu_a(X)\Big]$$

$$\text{Or,} \qquad R(\pi; w, \rho) := V(\pi; w) - V(\rho; w)$$

- Nice theories are obtained almost for free with fixed w and ρ.
- The influence function of $R(\pi; w, \rho)$ is given as

$$\psi(x, a, y; \phi, \mu) - R(\pi; w, \rho), \qquad \text{Lemma 2.3}$$

where $\psi(x, a, y; \phi_0, \mu_0) =$

$$w(x)\Bigg(\underbrace{\sum_{a'}\Big(\pi(a'|x) - \rho(a'|x)\Big)\mu(a'|x)}_{\text{main component}} + \underbrace{\frac{\pi-\rho}{\phi}(y-\mu)}_{\text{Doubly robust component}} .\Bigg)$$

(The arguments of π, ρ, ϕ, μ are omitted.)

- If the value (R) is estimated so that $\hat{R}_n = \mathbb{P}_n\psi + o_P(n^{-1/2})$, efficiency ($\sqrt{n}$-rate) and normality are obtained. (Section 2.2)

Finding the optimal weight and rule

A well-specified policy class Π_0

$$R(\pi; w, \rho) := \mathbb{E}\Big[w(X) \sum_{a\in\{0,1\}} \Big\{\pi(a|X) - \rho(a|X)\Big\}\mu_a(X)\Big] \quad \text{Population}$$

$$R_n(\pi; w, \rho) := \mathbb{P}_n\Big[w(X) \sum_{a\in\{0,1\}} \Big\{\pi(a|X) - \rho(a|X)\Big\}\mu_a(X)\Big] \quad \text{Sample}$$

- Does the optimal solution exist?
- Lemmas 2.1 and 2.2:
  If the true optimal rule is a restricted rule (or more generally, $\Pi^* \cap \Pi_0 \neq \emptyset$) before weighting and centering, it is true after weighting and centering. (Section 2.1)

Notation: * = optimal, $_0$ = restricted. E.g., Π_0^* = set of optimal restricted regimes.

Lemma 2.1. Suppose $\Pi^* \cap \Pi_0 \neq \varnothing$. Then $\Pi_0^*(w, \rho) = \Pi^* \cap \Pi_0$ for every $w \in \mathbb{R}_{++}^{\mathcal{X}}$, $\rho \in \mathbb{R}^{\mathcal{A}\times\mathcal{X}}$. In particular, if π is a solution to Equation (3) then $V(\pi) = V^*$.

Variance decomposition

Recall $\mathbb{P}_n\psi(\equiv \tilde{R}_n)$ is an efficient estimator.

$$\begin{aligned}\text{var}(\tilde{R}_n) &= \text{var}(R_n) + \text{var}(\tilde{R}_n - R_n) \\ &= \frac{1}{n}\underbrace{\text{var}\left(w\sum_a(\pi-\rho)\mu\right)}_{=\text{equation (5)}} + \frac{1}{n}\underbrace{\mathbb{E}\left[w^2\sum_a\frac{(\pi-\rho)^2}{\phi^2}\sigma^2\right]}_{=\text{equation (6)}}.\end{aligned}$$

(* The argument X or $(a|X)$ in w, π, ρ, μ_a, and σ_a is omitted for presentation.)

- Recall IF (ψ) is composed of the "main" and the "DR" components.
- (5) is the variance of value due to X-variation in the population (Notice the presence of μ_a).
- (6) quantifies any lack of overlap (Notice the presence of ϕ).
- Therefore, we want to find w and ρ that minimize (6).

Optimal retargeting–A minimax approach

The uniform (over-policies) objective

$$\text{Minimize } \Omega(w,\rho) \equiv \sup_{\pi\in\Pi} \mathbb{E}\left[w^2(X) \sum_a \frac{(\pi(a|X)-\rho(a|X))^2}{\phi(a|X)^2}\sigma^2(a|X)\right]$$

- Once we find the optimal (w_0, ρ_0), we find the optimal policy $\pi_* = \arg\max_\pi \hat{R}_n(\pi, w_0, \rho_0)$.
- $\Omega(w,\rho)$ controls "the uniform deviation between the estimated objective $(\tilde{V}_n)$ and the ideal finite-sample objective (V_n)"

Lemma 3.1. Suppose that $\phi(A \mid X)$ is bounded away from zero and Y is bounded. Let $w \in \mathbb{R}^{\mathcal{X}}_{++}$ with $\|w\|_\infty < \infty$ be given. Then there exists a universal constant C such that, for any $\delta \in (0, 1/2)$ and $\rho \in \mathbb{R}^{\mathcal{A}\times X}$, with probability at least $1-\delta$,

$$\sup_{\pi,\pi'\in\Pi_0} \left|(\tilde{V}_n(\pi;w) - \tilde{V}_n(\pi';w)) - (V_n(\pi;w) - V_n(\pi';w))\right|$$

- $$\leq C\left(\kappa(\Pi_0) + 1 + \sqrt{\log(1/\delta)}\right)\sqrt{\frac{\Omega(w,\rho)}{n}} + o\left(\frac{\log(1/\delta)}{\sqrt{n}}\right).$$

($\kappa(\cdot)$ = entropy integral)

Optimal retargeting–The solution for binary actions

The binary-action cases

Lemma 3.3. (ρ_0, w_0) minimizes $\Omega(w, \rho)$, where

$$\rho_0(+|x) = 1/2, \;\; w_0(x) \propto \left(\frac{\sigma^2(+|x)}{1+\phi(x)} + \frac{\sigma^2(-|x)}{1-\phi(x)} \right)^{-1},$$

and $\phi(x)$ is effect-coded.

- $\rho_0 = 1/2$ is a coin-flip rule, not depending on w.
- Assuming $\sigma(\cdot|x) = 1$, $w_0(x) \propto 1 - \phi^2(x)$, giving zero-weights to the no-overlaps.

Optimal retargeting–The solution for multi-arms

The multiple-action cases

Lemma 3.4. (ρ_0, w_0) minimizes $\Omega(w, \rho)$, where

$$\rho_0(a|x) = \frac{1}{2}\left(1 - \frac{\phi(a|x)}{\sigma^2(a|x)}\xi(x)\right)$$

$$w_0(x) \propto \left(\sum_{a\in\mathcal{A}} \frac{\sigma^2(a|x)}{\phi(a|x)} + \xi(x)/2\right)^{-1},$$

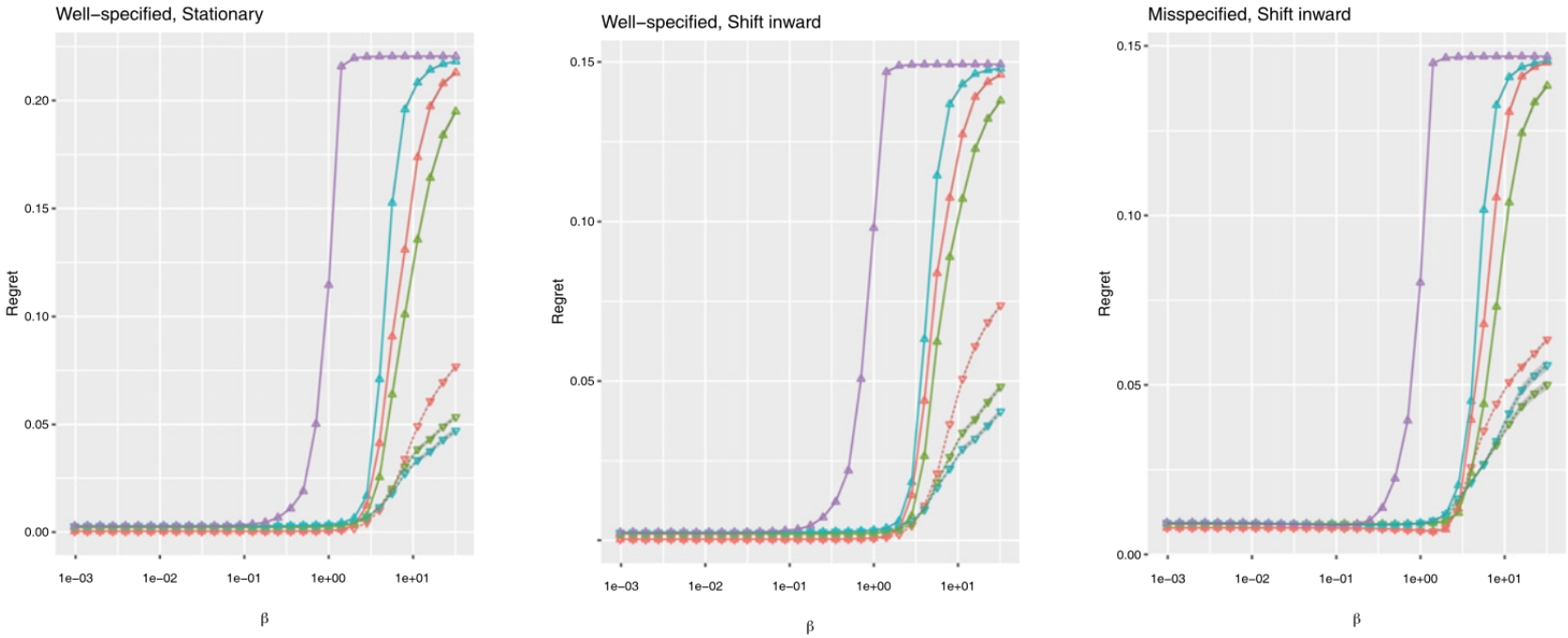
where $\xi(x) = (|\mathcal{A}| - 2)\left(\sum_{a\in\mathcal{A}} \frac{\phi(a|x)}{\sigma^2(a|x)}\right)^{-1}$.

- Now, ρ_0 depends on the choice of w when $|\mathcal{A}| > 2$.

Empirical evidences

Simulation study–Binary-action cases

Simulation ($|\mathcal{A}| = 2$, $n = 10,000$)



β controls the degree of overlap from good to bad.

- Works well under misspecification Π_0
- Works well under the train-test shift, $f_{X,\text{train}} \neq f_{X,\text{test}}$.

Job counseling–Multi-action cases

Personalized job counseling (Multi-Action)

Table 1. Average policy values of different policy learning methods applied to the job counseling dataset, with standard errors over replications.

	Average policy value (in 1000's)		
Method	Conventional	Retargeted	Improvement
DC	-3.44 ± 0.005	–	–
DM	-3.42 ± 0.005	-3.42 ± 0.005	0%
DR	1.14 ± 0.006	1.93 ± 0.006	70%
IPW	2.09 ± 0.006	2.65 ± 0.005	27%

Take-home message

Take-home message

- Lack of overlap $\Rightarrow$ efficiency loss (high variability)
- Frequently seen in observational data.
- If well specified, retargeting still finds the optimal rule (Lemmas 2.1, 2.2)
- Optimal weights can be found using a minimax approach

$$\text{Minimize } \Omega(w,\rho) \equiv \sup_{\pi\in\Pi} \mathbb{E}\left[w^2(X)\sum_a \frac{(\pi(a|X)-\rho(a|X))^2}{\phi(a|X)^2}\sigma^2(a|X)\right]$$

Lemmas 3.3 and 3.4 provide the solutions.

- Retargeting provides robust solutions against misspecification and train-test shift