Policy Evaluation with Latent Confounders via Optimal Balance

Andrew Bennett[1]
Cornell University
awb222@cornell.edu

Nathan Kallus[1]
Cornell University
kallus@cornell.edu

[1]Alphabetical order.

Policy Learning Problem

- Given some observational data on individuals described by some covariates (X), interventions performed on those individuals (T), and resultant outcomes (Y), wish to estimate utility of policies that assign treatment to individuals based on covariates
- Challenging problem when the relationship between T and Y in the logged data is confounded, even controlling for X
- Examples:
  - Drug assignment policy: X is patient information available to doctors, T is drug assigned, Y is medical outcome, and confounding due to factors not fully accounted for by X (e.g. socieoconomics) deciding drug assignment in observational data
  - Personalized education: X contains individual student statistics, T is an educational intervention, Y is measure of post-intervention student outcomes, and confounding due to X poorly accounting for criteria used by decision makers in observational data (e.g. X contains standardized test score but decisions made based on actual student capability)

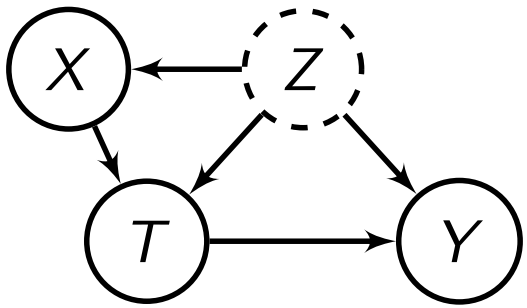
Setup - Latent Confounder Framework

Logged Data Model:

- Latent Confounders: $Z \in \mathcal{Z} \subseteq \mathbb{R}^p$
- Observed Proxies: $X \in \mathcal{X} \subseteq \mathbb{R}^q$
- Treatment: $T \in \{1, \ldots, m\}$
- Potential Outcomes: $Y(t) \in \mathbb{R}$

Assumption (Z are true confounders)

For every $t \in \{1, \ldots, m\}$, the variables $X, T, Y(t)$ are mutually independent, conditioned on Z.



Setup - Logging and Behavior Policies

Evaluation Policy:

- $\pi_t(x)$ denotes the probability of assigning treatment $T = t$ given observed proxies $X = x$ by *evaluation policy*

Logging Policy:

- $e_t(z)$ denotes the probability of assigning treatment $T = t$ given observed proxies $Z = z$ by *logging policy*
- $\eta_t(x)$ denotes the probability of assigning treatment $T = t$ given observed proxies $X = x$ by *logging policy*

Setup - Policy Evaluation Goal

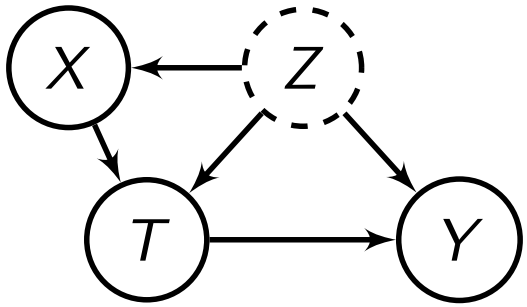
Definition (Policy Value)

$\tau^\pi = \mathbb{E}[\sum_{t=1}^m \pi_t(X) Y(t)]$.

Goal:

- Our goal is to estimate the policy value τ^π given iid logged data of the form $((X_1, T_1, Y_1), \ldots, (X_n, T_n, Y_n))$
- Want to find an estimator $\hat{\tau}^\pi$ that minimizes the MSE $\mathbb{E}[(\hat{\tau}^\pi - \tau^\pi)^2]$

Setup - Latent Confounder Model

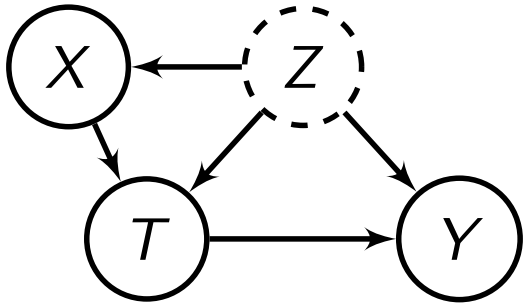


- We denote by $\varphi(z; x, t)$ the conditional density of Z given $X = x, T = t$

Assumption (Latent Confounder Model)

We assume that we have an identified model for $\varphi(z; x, t)$, and that we can calculate conditional densities and sample Z values using this model

Setup - Observed Proxies



- We do *not* assume ignorability given X
- This means standard approaches based on inverse propensity scores are bound to fail
- Instead the proxies X can be used (along with T) to calculate the posterior of the true confounders Z, which can be used for evaluation

Setup - Additional Assumptions

Assumption (Weak Overlap)

$\mathbb{E}[e_t^{-2}(Z)] < \infty$

Assumption (Bounded Variance)

The conditional variance of our potential outcomes given X, T is bounded: $\mathbb{V}[Y(t) \mid X, T] \leq \sigma^2$.

Setup - Mean Value Functions

Define the following mean value functions:

$$\mu_t(z) = \mathbb{E}[Y(t) \mid Z = z]$$
$$\nu_t(x, t') = \mathbb{E}[Y(t) \mid X = x, T = t'] = \mathbb{E}[\mu_t(Z) \mid X = x, T = t']$$
$$\rho_t(x) = \mathbb{E}[Y(t) \mid X = x] = \mathbb{E}[\mu_t(Z) \mid X = x]$$

Note that we can equivalently redefine policy value as:

$$\begin{aligned}\tau^{\pi} &= \mathbb{E}[\sum_{t=1}^{m} \pi_t(X) Y(t)] \\ &= \mathbb{E}[\sum_{t=1}^{m} \pi_t(X) \mu_t(Z)] \\ &= \mathbb{E}[\sum_{t=1}^{m} \pi_t(X) \nu_t(X, T)]\end{aligned}$$

Past Work - Standard Estimator Types

Weighted, Direct, and Doubly Robust estimators:

$$\hat{\tau}^{\pi}_{W} = \frac{1}{n}\sum_{i=1}^{n} W_i Y_i$$

$$\hat{\tau}^{\pi}_{\hat{\rho}} = \frac{1}{n}\sum_{i=1}^{n}\sum_{t=1}^{m} \pi_t(X_i)\hat{\rho}_t(X_i)$$

$$\hat{\tau}^{\pi}_{W,\hat{\rho}} = \frac{1}{n}\sum_{i=1}^{n}\sum_{t=1}^{m} \pi_t(X_i)\hat{\rho}_t(X_i) + \frac{1}{n}\sum_{i=1}^{n} W_i(Y_i - \hat{\rho}_{T_i}(X_i))$$

- Note that $\hat{\rho}_t$ is not straightforward to estimate via regression since $\rho_t(x) = \mathbb{E}[Y(t) \mid X = x] \neq \mathbb{E}[Y \mid X = x]$
- Correct IPW weights $W_i = \pi_{T_i}(X_i)/e_{T_i}(Z_i)$ are infeasible since Z_i is not observed, and naively misspecified IPW weights $W_i = \pi_{T_i}(X_i)/\eta_{T_i}(X_i)$ lead to biased evaluation

Past Work - Optimal Balancing

- Optimal Balancing (Kallus 2018) seeks to come up with a set of weights W_i that $\hat{\tau}^{\pi}_{W}$ minimize an estimate of the worst-case MSE of policy evaluation, given a class of functions for the unknown mean value function
- Define $CMSE(W, \mu)$ to be the conditional mean squared error given the logged data of $\hat{\tau}^{\pi}_{W}$ as an estimate of the sample average policy effect (SAPE), if the mean value function were given by μ
- Choose weights W^* for evaluation according to the rule:

$$W^* = \underset{W \in \mathcal{W}}{\arg\min} \sup_{\mu \in \mathcal{F}} CMSE(W, \mu)$$

- Permits simple QP algorithm when $\mathcal{F}$ is a class of RKHS functions

Generalized IPS Weights I

- Suppose we want to define weights $W(X, T)$ IPS-style such that the weighted estimator is unbiased term-by-term, this requires solving:

$$\mathbb{E}[W(X, T)\delta_{T_i t} Y(t)] = \mathbb{E}[\pi_t(X) Y(t)]$$

- Can easily verify that if we assume ignorability given X this equation is solved by standard IPS weights $W(X, T) = \pi_T(X)/\eta_T(X)$

Theorem (Generalized IPS Weights)

If $W(x, t)$ satisfies the above equation then for each $t \in \{1, \ldots, m\}$

$$W(x, t) = \pi_t(x) \frac{\sum_{t'=1}^{m} \eta_{t'}(x)\nu_t(x, t') + \Omega_t(x)}{\eta_t(x)\nu_t(x, t)},$$

for some $\Omega_t(x)$ such that $\mathbb{E}[\Omega_t(X)] = 0 \, \forall t$.

Generalized IPS Weights II

- Calculating these generalized IPS weights is not straightforward since it involves the counterfactual estimation of $\nu_t(x, t')$ for $t \neq t'$ (requires knowledge of Z)
- In addition would expect high variance from error in estimating ν_t due to its position in denominator
- However the fact that such weights exist supports idea of using optimal balancing style approach, and choosing weights that balance a flexible class of possible mean outcome functions

Adversarial Objective Motivation

Define the following, where we embed the dependence on μ inside ν_t implicitly:

$$f_{it} = W_i \delta_{T_i t} - \pi_t(X_i)$$

$$J(W, \mu) = \left(\frac{1}{n} \sum_{i=1}^{n} \sum_{t=1}^{m} f_{it} \nu_t(X_i, T_i) \right)^2 + \frac{2\sigma^2}{n^2} \|W\|_2^2,$$

Theorem (CMSE Upper Bound)

$\mathbb{E}[(\hat{\tau}_W^{\pi} - \tau^{\pi})^2 \mid X_{1:n}, T_{1:n}] \leq 2J(W, \mu) + O_p(1/n).$

Lemma (CMSE Convergence implies Consistency)

If $\mathbb{E}[(\hat{\tau}_W^{\pi} - \tau^{\pi})^2 \mid X_{1:n}, T_{1:n}] = O_p(1/n)$ *then* $\hat{\tau}_W^{\pi} = \tau^{\pi} + O_p(1/\sqrt{n})$.

Balancing Objective

Our optimal balancing objective is to choose weights W^* for evaluation according to the following optimzation problem:

$$W^* = \underset{W \in \mathcal{W}}{\arg\min} \sup_{\mu \in \mathcal{F}} J(W, \mu)$$

Feasibility of Balancing Objective I

- Minimizing $J(W, \mu)$ over some class of $\mu \in \mathcal{F}$ corresponds to balancing some class of functions ν implicitly indexed by μ, since:

$$J(W, \mu) = \left(\frac{1}{n} \sum_{i=1}^{n} W_i \nu_{T_i}(X_i, T_i) - \frac{1}{n} \sum_{i=1}^{n} \sum_{t=1}^{m} \pi_t(X_i) \nu_t(X_i, T_i) \right)^2 + \frac{2\sigma^2}{n^2} \|W\|_2^2$$

- Note that such balancing would be impossible over a generic flexible class of functions ν ignoring Z, due to $\nu_t(x, t')$ terms for $t \neq t'$

Feasibility of Balancing Objective II

- The following lemma suggests that this fundamental counterfactual issue may not be a problem given our implicit constraint imposed by indexing using μ and our overlap assumption:

Lemma (Mean Value Function Overlap)

Assuming $\|\mu_t\|_\infty \leq b$, under our weak overlap assumption, for all $x \in \mathcal{X}$, and $t, t', t'' \in \{1, \ldots, m\}$ we have

$$|\nu_t(x, t'')| \leq \frac{\eta_{t'}(x)}{\eta_{t''}(x)} \sqrt{8b\mathbb{E}[e_t^{-2}(Z) \mid X = x, T = t']|\nu_t(x, t')|}.$$

Assumptions for Consistent Evaluation I

Define $\mathcal{F}_t = \{\mu_t : \exists(\mu'_1, \ldots, \mu'_m) \in \mathcal{F} \text{ with } \mu'_t = \mu_t\}$, then we make the following assumptions:

Assumption (Normed)

For each $t \in \{1, \ldots, m\}$ *there exists a norm* $\|\cdot\|_t$ *on* $\text{span}(\mathcal{F}_t)$*, and there exists a norm* $\|\cdot\|$ *on* $\text{span}(\mathcal{F})$ *which is defined given some* $\mathbb{R}^m$ *norm as* $\|\mu\| = \|(\|\mu_1\|_1, \ldots, \|\mu_m\|_m)\|$.

Assumption (Absolutely Star Shaped)

For every $\mu \in \mathcal{F}$ *and* $|\lambda| \leq 1$*, we have* $\lambda\mu \in \mathcal{F}$.

Assumption (Convex Compact)

$\mathcal{F}$ *is convex and compact*

Assumptions for Consistent Evaluation II

Assumption (Square Integrable)

For each $t \in \{1, \ldots, m\}$ *the space* $\mathcal{F}_t$ *is a subset of* $\mathcal{L}^2(\mathcal{Z})$*, and its norm dominates the* $\mathcal{L}^2$ *norm (i.e.,* $\inf_{\mu_t \in \mathcal{F}_t} \|\mu_t\| / \|\mu_t\|_{\mathcal{L}^2} > 0$*).*

Assumption (Nondegeneracy)

Define $\mathcal{B}(\gamma) = \{\mu \in \mathsf{span}(\mathcal{F}) : \|\mu\| \leq \gamma\}$*. Then we have* $\mathcal{B}(\gamma) \subseteq \mathcal{F}$ *for some* $\gamma > 0$*.*

Assumption (Boundedness)

$\sup_{\mu \in \mathcal{F}} \|\mu\|_\infty < \infty$*.*

Assumptions for Consistent Evaluation III

Definition (Rademacher Complexity)

$\mathcal{R}_n(\mathcal{F}) = \mathbb{E}[\sup_{f \in \mathcal{F}} \frac{1}{n} \sum_{i=1}^{n} \epsilon_i f(Z_i)]$, where ϵ_i are iid Rademacher random variables.

Assumption (Complexity)

For each $t \in \{1 \ldots, m\}$ *we have* $\mathcal{R}_n(\mathcal{F}_t) = o(1)$.

Optimization Problem Convergence

Lemma (Minimax Lemma)

Let $B(W, \mu) = \frac{1}{n} \sum_{i=1}^{n} \sum_{t=1}^{m} f_{it} \nu_t(X_i, T_i)$. *Then under our consistency assumptions for every* $M > 0$ *we have the bound*

$$\min_{W} \sup_{\mu \in \mathcal{F}} J(W, \mu) \leq \sup_{\mu \in \mathcal{F}} \min_{\|W\|_2 \leq M} B(W, \mu)^2 + \frac{\sigma^2}{n^2} M^2.$$

Lemma (Optimization Problem Convergence)

Under our consistency assumptions we have
$\inf_W \sup_{\mu \in \mathcal{F}} J(W, \mu) = O_p(1/n)$.

Convergence Proof Sketch

- First, Minimax Lemma tells us that it is sufficient to prove $O_p(1/n)$ bound by picking a W in response to each possible μ such that:
  - $B(W(\mu), \mu) = 0 \forall \mu$
  - $\sup_{\mu \in \mathcal{F}} \|W(\mu)\|^2 = O_p(\sqrt{n})$
- Choose $W(\mu)$ as solution to: $\arg\min_W W^2$ s.t. $B(W, \mu) = 0$
- By Lagrangian Duality can find closed form solution to this problem, and prove $O_p(\sqrt{n})$ bound for solution using empirical process arguments and previous Mean Value Function Overlap lemma

Consistent Evaluation Theorem

Theorem (Root-n Consistency)

Under our consistency assumptions and assuming that $\mu \in \mathcal{F}$ we have $\hat{\tau}^{\pi}_{W^*} = \tau^{\pi} + O_p(1/\sqrt{n})$.

Proof idea:

- Define W^* as solution to $\inf_W \sup_{\mu \in \mathcal{F}} J(W, \mu)$
- Then assuming $\mu \in \mathcal{F}$ it must be case that $J(W^*, \mu) = O_p(1/n)$
- Given this $\sqrt{n}$ consistency follows automatically from previous theorems and lemmas

RKHS Class for Policy Evaluation I

Definition (Kernel Class)

$\mathcal{F}^K = \{\mu : ||\mu|| \leq 1\}$, where $||(\mu_1, \ldots, \mu_m)|| = \sqrt{\sum_{t=1}^m ||\mu_t||_K^2}$.

Theorem (Root-n Consistent Evaluation with Kernel Class)

Assuming K is a Mercer kernel (continuous and positive definite) and is bounded, $\mathcal{F}^K$ satisfies our assumptions for consistency.

Note that these assumptions are easily met for instance by the commonly used Gaussian kernel

RKHS Class for Policy Evaluation II

- Note that $\mathcal{F}^K$ having maximum norm 1 is without loss of generality, because if we wanted the maximum norm to instead be γ we could replace the Σ matrix by $\Gamma = \frac{1}{\gamma}\Sigma$ in our objective function, resulting in an equivalent re-scaled optimization problem
- Therefore we replace the Σ matrix in the objective with Γ, which is treated as a regularization hyperparameter

Kernel Balancing Algorithm I

Theorem

Define $Q_{ij} = \mathbb{E}[K(Z_i, Z_j')]$, $G_{ij} = \frac{1}{n^2}(Q_{ij}\delta_{T_i T_j} + \Gamma_{ij})$, *and* $a_i = \frac{2}{n^2}\sum_{j=1}^{n} Q_{ij}\pi_{T_j}(X_i)$, *where for each* i Z_i *and* Z_i' *are iid shadow variables. Then for some* c *that is constant in* W *we have the identity*

$$\sup_{\mu \in \mathcal{F}^K} J(W, \mu) = W^T G W - a^T W + c.$$

- Note that this means we can calculate our weights for consistent policy evaluation by solving a QP
- We can estimate Q given our assumption that we have an identified model for the posterior $\varphi(z; x, t)$

Kernel Balancing Algorithm II

Algorithm 1 Optimal Kernel Balancing

Input: Data $(X_{1:n}, T_{1:n})$, policy π, kernel function K, posterior density φ, regularization matrix Γ, number samples B, optional weight space $\mathcal{W}$ (defaults to $\mathbb{R}^n$ if not provided)

Output: Optimal balancing weights $W_{1:n}$

1: **for** $i \in \{1, \ldots, n\}$ **do**

2: **Sample Data.** Draw B data points Z_i^b from the posterior $\varphi(\cdot\,; X_i, T_i)$

3: **end for**

4: **Estimate Q.** Calculate $Q_{ij} = \frac{1}{B^2} \sum_{b=1}^{B} \sum_{c=1}^{B} K(Z_i^b, Z_i^c)$

5: **Calculate QP Inputs.** Calculate $G_{ij} = Q_{ij} \delta_{T_i T_j} + \Gamma_{ij}$, and $a_i = 2 \sum_{j=1}^{n} Q_{ij} \pi_{T_j}(X_i)$

6: **Solve Quadratic Program.** Calculate $W = \arg\min_{W \in \mathcal{W}} W^T G W - a^T W$

Experiment Setup - Data Generating Process and Policy

- Assume the following GLM-style data generating process:

$$Z \sim \mathcal{N}(0,1) \qquad X \sim \mathcal{N}(\alpha^T Z + \alpha_0, \sigma_X^2)$$
$$P_T = \beta^T Z + \beta_0 \qquad T \sim \text{softmax}(P_T)$$
$$W(t) \sim \mathcal{N}(\zeta(t)^T Z + \zeta_0(t), \sigma_Y^2) \qquad Y(t) = g(W(t))$$

- We assume Z is 1-dimensional, X is 10-dimensional, and use 2 treatment levels
- We experiment with the following functions for g:
  step : $g(w) = 3\mathbb{1}_{\{w \geq 0\}} - 6$
  exp : $g(w) = \exp(w)$
  cubic : $g(w) = w^3$
  linear : $g(w) = w$
- We experiment with evaluating the following parameterized policy:

$$\pi_t(X) = \frac{\exp(\psi_t^T X)}{\exp(\psi_1^T X) + \exp(\psi_2^T X)}$$

Experiment Setup - Method and Baselines

We experiment with the following methods in our evaluation:

- **OptZ** Our method, using $\Gamma = \gamma\,\mathsf{Identity}(n)$ for $\gamma \in \{0.001, 0.2, 1.0, 5.0\}$
- **IPS** IPS weights based on X using estimated $\hat{\eta}_t$
- **OptX** The optimal weighting method of (Kallus 2018) with same values of Γ as our method
- **DirX** Direct method by fitting $\hat{\rho}_t(x)$ incorrectly assuming ignorability given X
- **DirZ** Direct method by first fitting $\hat{\mu}_t$ using posterior samples from φ, then using the estimate $\hat{\rho}_t(x) = (1/D)\sum_{i=1}^{D} \hat{\mu}_t(z_i')$, where z_i' are sampled from $\varphi(\cdot; x, t)$
- **D:W** Doubly robust estimation using direct estimator **D** and weighted estimator **W**

Experiment Results - RMSE Convergence

n	$\mathbf{OptZ}_{0.001}$	$\mathbf{OptZ}_{0.2}$	$\mathbf{OptZ}_{1.0}$	$\mathbf{OptZ}_{5.0}$
200	$.39 \pm .07$	$.24 \pm .02$	$.36 \pm .02$	$.81 \pm .02$
500	$.19 \pm .02$	$.18 \pm .02$	$.23 \pm .02$	$.49 \pm .02$
1000	$.11 \pm .01$	$.11 \pm .01$	$.13 \pm .01$	$.27 \pm .01$
2000	$.08 \pm .01$	$.08 \pm .01$	$.09 \pm .01$	$.17 \pm .01$

n	**DirX**	**DirZ**	$\mathbf{DirX{:}OptZ}_{0.001}$	$\mathbf{DirZ{:}OptZ}_{0.001}$
200	$.52 \pm .02$	$2.6 \pm .02$	$.57 \pm .06$	$.41 \pm .07$
500	$.48 \pm .02$	$2.6 \pm .01$	$.55 \pm .03$	$.20 \pm .02$
1000	$.39 \pm .02$	$2.0 \pm .01$	$.49 \pm .02$	$.11 \pm .01$
2000	$.40 \pm .01$	$2.0 \pm .01$	$.48 \pm .01$	$.08 \pm .01$

n	**IPS**	$\mathbf{OptX}_{0.001}$	$\mathbf{OptX}_{0.2}$	$\mathbf{OptX}_{1.0}$	$\mathbf{OptX}_{5.0}$
200	$.47 \pm .03$	$2.0 \pm .03$	$2.1 \pm .03$	$2.3 \pm .02$	$2.5 \pm .02$
500	$.48 \pm .03$	$2.0 \pm .02$	$2.1 \pm .02$	$2.3 \pm .02$	$2.6 \pm .02$
1000	$.39 \pm .02$	$2.0 \pm .01$	$2.1 \pm .01$	$2.3 \pm .01$	$2.5 \pm .01$
2000	$.40 \pm .01$	$2.0 \pm .01$	$2.1 \pm .01$	$2.3 \pm .01$	$2.5 \pm .01$

Experiment Results - Bias Convergence

n	$\mathbf{OptZ}_{0.001}$	$\mathbf{OptZ}_{0.2}$	$\mathbf{OptZ}_{1.0}$	$\mathbf{OptZ}_{5.0}$
200	$.03 \pm .39$	$.11 \pm .21$	$.29 \pm .21$	$.78 \pm .18$
500	$.09 \pm .17$	$.10 \pm .15$	$.17 \pm .16$	$.47 \pm .15$
1000	$.02 \pm .11$	$.05 \pm .09$	$.08 \pm .09$	$.25 \pm .09$
2000	$.03 \pm .07$	$.05 \pm .06$	$.07 \pm .07$	$.16 \pm .07$

n	**DirX**	**DirZ**	$\mathbf{DirX:OptZ}_{0.001}$	$\mathbf{DirZ:OptZ}_{0.001}$
200	$.49 \pm .18$	$2.6 \pm .14$	$.43 \pm .38$	$.05 \pm .40$
500	$.45 \pm .16$	$2.6 \pm .12$	$.51 \pm .19$	$.10 \pm .18$
1000	$.46 \pm .15$	$2.6 \pm .11$	$.47 \pm .13$	$.04 \pm .11$
2000	$.42 \pm .17$	$2.6 \pm .11$	$.47 \pm .09$	$.03 \pm .07$

n	**IPS**	$\mathbf{OptX}_{0.001}$	$\mathbf{OptX}_{0.2}$	$\mathbf{OptX}_{1.0}$	$\mathbf{OptX}_{5.0}$
200	$.40 \pm .25$	$1.9 \pm .21$	$2.1 \pm .20$	$2.3 \pm .19$	$2.5 \pm .18$
500	$.43 \pm .21$	$2.0 \pm .16$	$2.1 \pm .15$	$2.3 \pm .14$	$2.6 \pm .13$
1000	$.37 \pm .12$	$2.0 \pm .10$	$2.1 \pm .09$	$2.3 \pm .09$	$2.5 \pm .08$
2000	$.39 \pm .10$	$2.0 \pm .08$	$2.1 \pm .07$	$2.3 \pm .07$	$2.5 \pm .07$

Experimental Results - Analysis

- Experimental Results seem to support our theory of consistency of our policy value estimator
- Standard baselines naively assuming ignorability given X were all biased
- Direct estimation was not consistent even when taking latent structure into account
- Doubly robust estimation did not help over simple weighted estimation

Possible Questions for Future Work

- How to perform inference on policy value estimates using our method?
- How to perform policy improvement using our method?
- Is there a better, consistent way to fit $\hat{\rho}_t$ for direct evaluation?
- How to optimize adversarial objective over different function classes (e.g. neural networks)?
- Can we establish semiparametric efficiency, or extend methodology to achieve semiparametric lower bound?
- Finite sample bounds for our method?
- How do we extend methodology to situation where we don't have an identified model for $\varphi(z; x, t)$?