Off-policy Evaluation Beyond Overlap: Sharp Partial Identification Under Smoothness

Samir Khan [1] **Martin Saveski** [2] **Johan Ugander** [3]

Abstract

Off-policy evaluation, and the complementary problem of policy learning, use historical data collected under a logging policy to estimate and/or optimize the value of a target policy. Methods for these tasks typically assume overlap between the target and logging policy, enabling solutions based on importance weighting and/or imputation. Absent such an overlap assumption, existing work either relies on a well-specified model or optimizes needlessly conservative bounds. In this work, we develop methods for no-overlap policy evaluation without a well-specified model, relying instead on non-parametric assumptions on the expected outcome, with a particular focus on Lipschitz smoothness. Under such assumptions we are able to provide sharp bounds on the off-policy value, along with asymptotically optimal estimators of those bounds. For Lipschitz smoothness, we construct a pair of linear programs that upper and lower bound the contribution of the no-overlap region to the off-policy value. We show that these programs have a concise closed form solution, and that their solutions converge under the Lipschitz assumption to the sharp partial identification bounds at a minimax optimal rate, up to log factors. We demonstrate the effectiveness our methods on two semi-synthetic examples, and obtain informative and valid bounds that are tighter than those possible without smoothness assumptions.

[1]Department of Statistics, Stanford University [2]Information School, University of Washington [3]Department of Management Science and Engineering, Stanford University. Correspondence to: Samir Khan <samirk@stanford.edu>.

Proceedings of the 41st International Conference on Machine Learning, Vienna, Austria. PMLR 235, 2024. Copyright 2024 by the author(s).

1. Introduction

Off-policy evaluation (OPE) is the task of estimating the value of an evaluation/target policy using data from a behavior/logging policy, and arises naturally in many settings (Li et al., 2010; 2011; Bottou et al., 2013; Swaminathan et al., 2017; Liao et al., 2021; Chin et al., 2022). The two standard approaches to off-policy evaluation are reweighting and imputation. Reweighting methods, as the name suggests, reweight outcomes observed under the behavior policy to obtain unbiased estimates of the evaluation policy. Imputation methods, on the other hand, model the expected outcome of taking an action as a function of covariates, and then use this model to estimate the off-policy value. Finally, doubly-robust methods combine these two approaches to obtain better theoretical guarantees (Dudík et al., 2011). Once an estimator of the off-policy value is available, the estimated policy value can be optimized over a policy class to select an optimal policy, which is the task of policy learning.

All of these methods for policy evaluation and learning generally rely on an overlap assumption, which ensures that any action with positive probability under the evaluation policy also has positive probability under the behavior policy. If the overlap assumption is not satisfied, the weights used by reweighting methods will be infinite, and the models used by imputation methods will only be valid under strong well-specification assumptions. In both cases, overlap violations lead to biased estimates of the off-policy value and sub-optimal choices of learned policy (Sachdeva et al., 2020).

This state of affairs raises the question: if we do not have overlap and are unwilling to assume a well-specified model, can we say anything about the off-policy value? In such a case, the off-policy value is not point-identified, so we cannot provide a point estimate, but can still provide partial identification bounds on the off-policy value.

In this work, we propose new estimators and resolve open questions in the policy evaluation and learning literatures by identifying tight bounds on the off-policy value in the presence of overlap violations under modest non-parametric assumptions on the expected outcome function, and give rate-optimal estimators of these bounds (up to log-factors). We focus on smoothness assumptions that constrain the

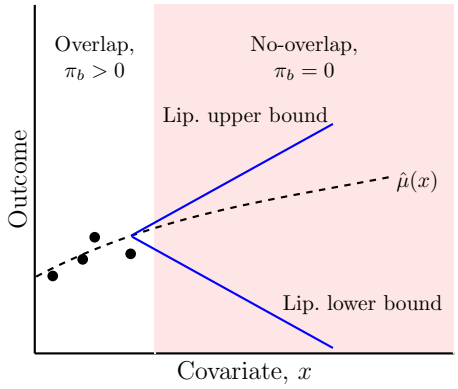

Figure 1. Visualization of our approach on a problem with four data points.

conditional outcome to be L-Lipschitz, and briefly discuss other assumptions such as monotonicity. These smoothness assumptions are generalizations of the classical bounded response assumption for partial identification proposed by Manski (1990).

The spirit of our approach is visualized in Figure 1, which shows a toy policy problem with a single real-valued covariate. For small values of the covariate, we have overlap and so the behavior policy has a positive probability of assigning units to treatment. For those treated units, we observe their outcomes (shown as black dots). On the other hand, for large values of the covariate, we do not have overlap (indicated by red), and so units in this region are never treated and we never observe any responses. We can fit a model $\hat{\mu}(x)$ (shown as a dashed black line) to the observed responses, but because this model has not been trained on observations in the no-overlap region, there is no way to guarantee that its predictions there are remotely accurate.

Rather than directly using a model $\hat{\mu}$, in this work we make the weaker assumption that the true conditional mean function μ is L-Lipschitz. Under this assumption, if the model $\hat{\mu}$ is consistent in the overlap region, we can give bounds (shown in blue) on the conditional mean in the no-overlap region. We emphasize that we assume smoothness with respect to a particular covariate space and metric; as such, the covariate space and metric should be chosen based on domain knowledge about which covariates and metric the outcome is plausibly smooth in.

2. Related Work

The overlap assumption is standard in off-policy evaluation literature (Bottou et al., 2013; Swaminathan and Joachims, 2015; Thomas and Brunskill, 2016; Wang et al., 2017) and we focus our discussion on the emerging literature on policy evaluation and learning in the no-overlap setting.

Well-specified outcome models. One strand of this literature relies on access to a well-specified outcome model. For example, Mou et al. (2023) assume that the reward function lies within a reproducing kernel Hilbert space and use this assumption to extrapolate into the no-overlap region, thus effectively assuming a well-specified outcome model. Additionally, their method requires the action probabilities to be positive, albeit potentially arbitrarily small, while our results allow for action probabilities that are exactly zero. On the policy learning side, Sachdeva et al. (2020) first show that a policy learned using inverse-propensity scoring is sub-optimal, and then consider several remedies based on restricting the set of policies that are optimized over or using a well-specified model. In contrast to these works, we require no such extrapolation and allow for evaluation of arbitrary policies.

Learning by optimizing lower bounds. Another strand of the no-overlap literature develops methods for policy learning by optimizing lower bounds on the off-policy value over some class of outcome functions. This is the approach taken by Higbee (2022) and Ben-Michael et al. (2021), although in both cases the bounds they optimize are potentially quite loose and are thus inappropriate for providing information about the value of a particular policy; we demonstrate this looseness in simulations in Section 6. Of these two works, Higbee (2022) is slightly further from ours, in that they make assumptions on the expected outcome as a function of action rather than as a function of covariates, and Ben-Michael et al. (2021) is slightly closer to ours, in that they make a similar Lipschitz assumption on the expected outcome as a function of covariates. In both cases, our estimators are different and the theoretical optimality results we obtain for our estimators go beyond the results found in these works and more completely develop the partial identification framework for off-policy evaluation under smoothness; see Appendix E for further details.

Reinforcement learning. In the reinforcement learning literature, Jiang and Huang (2020) propose a "minimax value interval" that is valid even under no-overlap. However, their method relies on a well-specified outcome model, requires solving a challenging min-max optimization problem, and is more relevant when both the outcome model and behavior policies are unknown. In our setting, the behavior policy is known, and so our methods, which also come with optimality guarantees, are preferable.

There are also proposals in the reinforcement learning literature to learn a policy by optimizing a lower bound, such as that of Xie et al. (2021). This work requires knowledge of a model class $\mathcal{F}$ that is known to contain good approximations of the true value function, and which can be taken to be the class of Lipschitz functions in our setting. There

are two differences between this work and ours. First, Xie et al. (2021) require access to an empirical risk minimization oracle over the class $\mathcal{F}$, which is computationally expensive when $\mathcal{F}$ is a non-parametric class like the class of Lipschitz functions. Thus the methods of Xie et al. (2021) are better suited to parametric $\mathcal{F}$.

Second, the lower bound that Xie et al. (2021) optimize is the worst-case off-policy value over value functions in $\mathcal{F}$ that have small sample error. Rather than working over this subset of $\mathcal{F}$, we impose an additional "consistency" assumption on $\mathcal{F}$ (see Section 4 for details) that ensures that the functions in $\mathcal{F}$ have zero population error, and then find worst-case bounds on the off-policy value over the entirety of $\mathcal{F}$. When the true value function is contained in $\mathcal{F}$ and the number of samples is large, the bounds obtained by our approach and their approach will be nearly equal.

Peer-review applications. Finally, Saveski et al. (2023) deploy methods based on partial identification under smoothness assumptions in an off-policy evaluation problem arising in the context of matching reviewers to papers at academic conferences. However, they combine the steps of fitting a model $\hat{\mu}$ and estimating bounds on the off-policy value into a single linear program, leading to sub-optimal bounds. Their work highlights the importance of the no-overlap setting, and thus of the potential for our methods and results to provide stronger conclusions on existing data sets.

3. Model and Notation

We consider the problem of off-policy evaluation of stochastic policies that select an action from an action space based on covariates X_i. The covariates X_i take values in a metric space $\mathcal{X}$ with metric $d(\cdot,\cdot)$. The actions are random variables A_i that takes values in a set $\mathcal{A}$. Each action $a \in \mathcal{A}$ has a corresponding potential outcome $Y_i(a)$, and we assume that tuples $(X_i, \{Y_i(a)\}_{a\in\mathcal{A}})$ are i.i.d. from a distribution P_0 on $\mathcal{X} \times \mathbb{R}^{|\mathcal{A}|}$ for some covariate space $\mathcal{X}$. The actions A_i are such that $\mathbb{P}(A_i = a \mid X_i = x) = \pi_b(x, a)$ for a behavior policy $\pi_b : \mathcal{X} \times \mathcal{A} \to [0, 1]$ that satisfies the constraint $\sum_{a\in\mathcal{A}} \pi_b(x, a) = 1$, for all $x \in \mathcal{X}$.

In this set-up, we observe $\{(X_i, A_i, Y_i(A_i))\}_{i=1}^n$ under the behavior policy π_b and would like to estimate the value of a different policy π_e. That is, we would like to estimate the functional $\psi(P_0) = \mathbb{E}_{(X_i,Y_i)\sim P_0,\ A_i|X_i\sim\pi_e}[Y_i(A_i)]$, where $A_i|X_i$ is drawn according to π_e. It is convenient to decompose this functional across the actions and write it as $\psi(P_0) = \sum_{a\in\mathcal{A}} \mathbb{E}_{(X_i,Y_i)\sim P_0}[Y_i(a)\pi_e(X_i, a)]$.

Critically, in this work we allow there to exist $x \in \mathcal{X}$ and $a \in \mathcal{A}$ such that $\pi_b(x, a) = 0$. We refer to $\{x : \pi_b(x, a) = 0\}$ as the *no-overlap region* and to $\{x : \pi_b(x, a) > 0\}$ as the *overlap region*.

The model described thus far corresponds to a general multi-action off-policy evaluation problem. However, by virtue of the decomposition across actions given above, we can naturally reduce any multi-action OPE problem to a binary action OPE problem. For any action $a \in \mathcal{A}$, we can define the binary action $\tilde{A}_i = \mathbf{1}\{A_i = a\}$ and binary evaluation policy $\tilde{\pi}_e(X_i) = \pi_e(X_i, a)$. Then, if we can estimate the functional $\psi(P_0) = \mathbb{E}_{P_0}[Y_i(a)\tilde{\pi}_e(X_i)]$, from data $(X_1, \tilde{A}_1, Y_1(a)\tilde{A}_1), \cdots, (X_n, \tilde{A}_n, Y_n(a)\tilde{A}_n)$, we can combine these estimates across all $a \in \mathcal{A}$ to estimate $\psi(P_0)$.

With this in mind, when developing our methods and theoretical results, we consider binary off-policy evaluation problems with action space $\mathcal{A} = \{0, 1\}$ and $Y_i(0) = 0$. Since there is only one action with non-zero reward, we suppress the dependence on the action in our notation, writing Y_i for $Y_i(1)$, $\pi_b(X_i)$ for $\pi_b(X_i, 1)$, and $\pi_e(X_i)$ for $\pi_e(X_i, 1)$. Further, in the binary problem, we write $\mu_P(x) = E_P[Y_i(1) \mid X_i = x]$ for the conditional mean function under a distribution P and $\hat{\mu}_P$ for an estimate of μ_P learned from the observed data through empirical risk minimization.

4. Nonparametric Partial Identification

In this section, we describe our framework for OPE without overlap and then provide several specific instantiations. We work in the setting where A_i is binary; by the reduction of Section 3, our results extend naturally to the multi-action setting. For an OPE problem with binary actions we show how to partially identify the off-policy value $\psi(P_0)$ using the assumption that $P_0 \in \mathcal{P}$ for a family of distributions $\mathcal{P}$. This family $\mathcal{P}$ encodes the nature of our assumptions on P_0 and may take several forms, but one crucial feature that $\mathcal{P}$ has to satisfy is that it must only contain distributions that are consistent with the true distribution P_0. This means that each $P \in \mathcal{P}$ must have the same marginal distribution of X_i as P_0, and the same joint distribution of (X_i, Y_i) in the overlap region as P_0.

We describe our approach for a generic family $\mathcal{P}$. The first step is to write $\psi = \psi_1 + \psi_2$ where

$$\begin{aligned} \psi_1(P) &= \mathbb{E}_P[Y_i\pi_e(X_i)\mathbf{1}\{\pi_b(X_i) > 0\}], \\ \psi_2(P) &= \mathbb{E}_P[Y_i\pi_e(X_i)\mathbf{1}\{\pi_b(X_i) = 0\}], \end{aligned} \tag{1}$$

so that ψ_1 is the contribution of the overlap region and ψ_2 is the contribution of the no-overlap region. The first term, ψ_1, is identifiable, so we can estimate it, e.g., using an inverse-probability weighted (IPW) estimator $\hat{\psi}_1$ (a self-normalized or doubly-robust estimator can also be used to estimate ψ_1 without any modifications to our results (Dudík et al., 2011; Swaminathan and Joachims, 2015)).

The second term, ψ_2, however is not identified. Our approach is to bound its contribution to (1) using the as-

sumption that $P_0 \in \mathcal{P}$ for some family $\mathcal{P}$. Under this assumption, the tightest possible bounds we could obtain are $\inf_{P\in\mathcal{P}} \psi_2(P)$ and $\sup_{P\in\mathcal{P}} \psi_2(P)$, which we denote by ψ_2^- and ψ_2^+ respectively. So to bound ψ_2, we must construct estimators $\hat{\psi}_2^-$ and $\hat{\psi}_2^+$ of ψ_2^- and ψ_2^+, respectively.

Once we have such estimators, we can set $\hat{\psi}^- = \hat{\psi}_1 + \hat{\psi}_2^-, \hat{\psi}^+ = \hat{\psi}_1 + \hat{\psi}_2^+$, and use $[\hat{\psi}^-, \hat{\psi}^+]$ as an interval estimate of $\psi(P_0)$. The following result, whose proof appears in Appendix A, guarantees the validity of this interval under conditions on $\hat{\psi}_1, \hat{\psi}_2^+$, and $\hat{\psi}_2^-$.

Theorem 4.1. *Suppose that $\hat{\psi}_1$ is a consistent estimator of $\psi_1(P_0)$, and that $\hat{\psi}_2^-$ and $\hat{\psi}_2^+$ are consistent estimators of ψ_2^- and ψ_2^+ respectively. Then, for any $\epsilon > 0$, $\lim_{n\to\infty} \mathbb{P}(\hat{\psi}^- - \epsilon \le \psi(P_0) \le \hat{\psi}^+ + \epsilon) = 1$.*

Thus, if we construct bounds $\hat{\psi}_2^-, \hat{\psi}_2^+$ satisfying the conditions of Theorem 4.1, the interval $[\hat{\psi}^-, \hat{\psi}^+]$ will be consistent for $\psi(P_0)$. Next, we consider specific choices of $\mathcal{P}$ that arise from different assumptions on P_0, and construct such an estimator $\hat{\psi}_2^-$. We focus throughout on the infimum, but all of our discussion holds *mutatis mutandis* for the supremum.

Boundedness assumptions. As a first example, we consider the following simple choice of $\mathcal{P}$, corresponding to the assumption that the response Y_i must lie in the interval $[\ell, u]$:

$$\mathcal{P}_{\ell,u}^{\text{bdd}} = \{P \text{ consistent w. } P_0 : \ell \le Y_i \le u \text{ a.s.}\}. \quad (2)$$

The assumption that $\ell \le Y_i \le u$ implies that $\ell \le \mu_P(x) \le u$ for all x as well, and a natural choice of ψ_2^- is $\hat{\psi}_2^- = \frac{\ell}{n}\sum_{i=1}^n \pi_e(X_i)\mathbf{1}\{\pi_b(X_i) = 0\}$. Then, by the law of large numbers, $\hat{\psi}_2^- \xrightarrow{\mathbb{P}} \ell\mathbb{E}[\pi_e(X_i)\mathbf{1}\{\pi_b(X_i) = 0\}]$, and we can verify that this is also the value of ψ_2^-. Thus the consistency conditions of Theorem 4.1 hold and the interval $[\hat{\psi}^-, \hat{\psi}^+]$ is consistent for $\psi(P_0)$.

The interval $[\hat{\psi}^-, \hat{\psi}^+]$ is an analogue of the so-called Manski bounds (Manski, 1990), and so our framework generalizes this well-established practice. As such, we can also obtain confidence intervals for the partial identification region of $\psi(P_0)$ using the methods of Imbens and Manski (2004) and Stoye (2009).

Smoothness assumptions. Next, we move on to our main focus: Lipschitz assumptions on $\mu_P(x)$. Formally, this corresponds to the family

$$\mathcal{P}_L^{\text{Lip}} = \{P \text{ consistent w. } P_0 : \mu_P \text{ is } L\text{-Lipschitz}\},$$

where by L-Lipschitz we mean that $|\mu_P(x_1) - \mu_P(x_2)| \le Ld(x_1, x_2)$ for some metric d on $\mathcal{X}$. By restricting μ_P to be L-Lipschitz, we can draw conclusions about the behavior

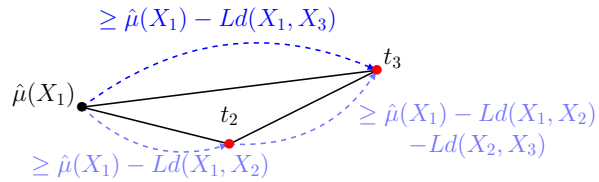

Figure 2. An illustration of the *no-interaction* property for Lipschitz constraints. At the optimal solution of (3), constraints (blue) between pairs of points in the no-overlap region (red) are not active (light blue).

of μ_P in the no-overlap region based on our observations in the overlap region and thus estimate $\inf_{P\in\mathcal{P}_L^{\text{Lip}}} \psi_2(P)$.

We propose to construct the estimator $\hat{\psi}_2^-$ in this setting by solving the following linear program:

$$\begin{aligned} \min_{t_1,\cdots,t_n} \quad & \frac{1}{n}\sum_{i=1}^n t_i \pi_e(X_i)\mathbf{1}\{\pi_b(X_i) = 0\} \\ \text{s.t.} \quad & |t_i - t_j| \le Ld(X_i, X_j), \quad 1 \le i < j \le n, \\ & t_i - \hat{\mu}(X_i) = 0, \quad \forall i \text{ s.t. } \pi_b(X_i) > 0. \end{aligned} \quad (3)$$

The problem (3) is an approximation of the population problem $\inf_{P\in\mathcal{P}_L^{\text{Lip}}} \psi_2(P)$ in three ways: *(i)* it averages over sample points in the objective rather than over P_0; *(ii)* it only enforces the Lipschitz constraint between pairs of observed data points X_i, X_j rather than between all pairs x_1, x_2; and *(iii)* it sets points in the overlap region to have value $\hat{\mu}$ rather than μ. We will see shortly that all of these approximations are asymptotically negligible.

We now characterize the solution to (3) and its properties under assumptions. The key to our results is the surprising fact that (3) can be solved in closed-form whenever d is a metric, even though this is not generally the case for linear programs. We are able to obtain a closed-form solution to (3) because its constraints satisfy what we refer to as a *no-interaction* property, by which we mean that points in the no-overlap region do not place sharp bounds on each other.

To see why, consider an example with $n = 3$, where the $i = 1$ point is in the overlap region and the other two points are in the no-overlap region (Figure 2). We must have $t_1 = \hat{\mu}(X_1)$. We must also have $t_2 \ge \hat{\mu}(X_1) - Ld(X_1, X_2)$, and since the objective is non-decreasing in the t_i, we set $t_2 = \hat{\mu}(X_1) - Ld(X_2, X_1)$. Then, consider t_3. The lower bound on t_3 coming directly from t_1 is $\hat{\mu}(X_1) - Ld(X_3, X_1)$, while the lower bound coming indirectly from t_2 is $\hat{\mu}(X_1) - Ld(X_2, X_1) - Ld(X_3, X_2)$.

The key point is that $d(X_2, X_1) + d(X_3, X_2) > d(X_3, X_1)$ by the triangle inequality, and so the bound from the overlap region is always sharper than the bound from the no-overlap

region. We emphasize that this no-interaction property is non-trivial and does not hold in general, e.g., it does not hold for an α-Hölder continuity assumption.

Based on this intuition, we expect that constraints between points in the no-overlap region in (3) are redundant, and the optimal solution to (3) will simply set each t_i in the no-overlap region to the tightest lower bound obtained from a point in the overlap region. We state this precisely in our next theorem, which requires the following assumptions.

Assumption 4.2. The estimated $\hat{\mu}$ satisfies $|\hat{\mu}(X_i) - \hat{\mu}(X_j)| \leq Ld(X_i, X_j)$ for all X_i, X_j such that $\pi_b(X_i), \pi_b(X_j) > 0$.

Assumption 4.3. The estimated $\hat{\mu}$ is consistent, so that $\sup_{x:\pi_b(x)>0} |\hat{\mu}(x) - \mu_{P_0}(x)| \xrightarrow{\mathbb{P}} 0$.

Assumption 4.2 is necessary for (3) to be feasible. To satisfy this assumption for moderate values of L, we recommend estimating $\hat{\mu}$ with smooth approximations such as parametric models, splines, or kernel methods (Hastie et al., 2009). Assumption 4.3 requires consistency in the overlap region; note that we make no assumptions on the behavior of $\hat{\mu}$ in the no-overlap region.

Our next assumption is on the marginal distribution $P_{0,X}$ of X_i. This assumption ensures that the marginal distribution of X_i does not have any "holes" that prevent us from observing parts of the overlap region.

Assumption 4.4. For every x such that $\pi_b(x) > 0$, either (a) the distribution $P_{0,X}$ has an atom at $\{x\}$ or (b) for every $\epsilon > 0$, there exists $\delta > 0$ such that $\mathbb{P}(d(X_i, x) \leq \epsilon) > \delta$.

Under these assumptions, we have the following result, whose proof appears in Appendix A. In this result, we write $\mathbf{1}_i$ as shorthand for $\mathbf{1}\{\pi_b(X_i) = 0\}$.

Theorem 4.5. *Suppose that $P_0 \in \mathcal{P}_L^{\mathrm{Lip}}$. Then, for $\hat{\psi}_2^-$ and ψ_2^-, we have that:*

(a) under Assumption 4.2, the problem (3) is feasible and has value $\hat{\psi}_2^-$, which is

$$\frac{1}{n}\sum_{i=1}^{n} \pi_e(X_i)\left(\max_{j:\pi_b(X_j)>0} \hat{\mu}(X_j) - Ld(X_i, X_j)\right)\mathbf{1}_i,$$

(b) the population bound is ψ_2^-, which is

$$\mathbb{E}_{P_0}\left[\pi_e(X_i)\left(\sup_{x:\pi_b(x)>0} \mu_{P_0}(x) - Ld(X_i, x)\right)\mathbf{1}_i\right],$$

(c) under Assumptions 4.3 and 4.4, we have $\hat{\psi}_2^- \xrightarrow{\mathbb{P}} \psi_2^-$.

Theorem 4.5(a) establishes a closed-form solution to (3), i.e., we do not need to solve it numerically. This is important even for moderate values of n, since (3) has $O(n^2)$ constraints. The fastest known algorithms for solving a general linear program with d constraints have complexity $O(d^{2.5})$ (Lee and Sidford, 2015), and so solving (3) has worst-case complexity $O(n^5)$. In contrast, computing the closed-form given in (12) requires only $O(n^2)$ operations, which is of the same order as computing all pairwise distances $d(X_i, X_j)$. In cases where even $O(n^2)$ operations are too expensive, we can construct conservative approximations of $\hat{\psi}_2^-$ by building on recent advances in efficient exact and approximate nearest neighbor search; we discuss such tools further in Appendix B.

The results in Theorem 4.5(b,c) are of a different flavor, and characterize the statistical properties of solutions to (3). Specifically, they show that three approximations we made in constructing $\hat{\psi}_2^-$ are asymptotically negligible, and we recover $\inf_{P\in\mathcal{P}_L^{\mathrm{Lip}}} \psi_2(P)$ in large samples. Thus the consistency conditions of Theorem 4.1 are satisfied and our bounds are the best possible under the given assumptions.

Further assumptions. The framework we present captures many other potentially interesting assumptions. For example, we can combine the two assumptions presented here, and assume that $P \in \mathcal{P}_L^{\mathrm{Lip}} \cap \mathcal{P}_{\ell,u}^{\mathrm{bdd}}$, to obtain tighter bounds than either assumption alone would give. In fact, the no-interaction property discussed after (3) continues to hold in this case, and our results then extend naturally; we present these more general results in Theorem A.1 of Appendix A.

A variety of other assumptions are also possible, including monotonicity of μ with respect to a partial order on the covariates X_i, convexity of μ, smoothness of higher derivatives of μ, α-Hölder continuity of μ, and compositions of these assumptions. Interestingly, these assumptions and their compositions do not necessarily satisfy the no-interaction property: for example, if we assume both Lipschitz smoothness and monotonicity, the no-interaction property no longer holds, as we show in Appendix D. As such, our results on smoothness, boundedness, and their composition, are both non-trivial and surprising.

In summary, we derive bounds on the off-policy value under smoothness or smoothness and boundedness assumptions without assuming overlap. Unlike previous work, our bounds are provably sharp—that is, they allow an analyst to make the strongest conclusions possible about a particular policy under the given assumptions—and can be computed efficiently even for large datasets owing to our careful analysis and closed-form solution of the linear programs involved.

5. Rates of Convergence

In this section, we precisely characterize the asymptotics of $\hat{\psi}_2^-$ by identifying the rate at which it converges to ψ_2^- and showing that this rate is optimal up to log factors. We first

upper bound the mean-squared error (MSE) of $\hat{\psi}_2^-$, with a proof given in Appendix A. Recall that we write $\mathbf{1}_i$ for $\mathbf{1}\{\pi_b(X_i) = 0\}$.

Theorem 5.1. *Let $\hat{\psi}_2^-, \psi_2^-$ be as in Theorem 4.5. Then, if X_i has a density that is lower bounded by a constant b then the mean-squared error $E[(\hat{\psi}_2^- - \psi_2^-)^2]$ is upper bounded by*

$$2\mathbb{E}[\|(\hat{\mu}(x) - \mu(x))\mathbf{1}_i\|_\infty^2] + 4L^2(c_d bn)^{-2/d} + \sigma^2 n^{-1}, \quad (4)$$

where c_d is the volume of a d-dimensional unit ball,

$$\sigma^2 = \mathrm{var}\left(\pi_e(X_i)[\sup_{x:\pi_b(x)>0} \mu_{P_0}(x) - Ld(X_i, x)]\mathbf{1}_i]\right).$$

The bound (4) contains three terms: one corresponding to error from the estimation of μ; one corresponding to the error from approximating a supremum by a maximum; and one that is noise.

Classical results on non-parametric regression show that the minimax rate for estimating an L-Lipschitz function in d dimensions in the sup-norm is of order $(\log n/n)^{-2/(d+2)}$, and that this rate is obtained by a kernel estimator (Stone, 1982). Thus, the dominant term of (4) is the first one, and the MSE of $\hat{\psi}_2^-$ is $O((\log n/n)^{-2/(d+2)})$ when using an appropriate kernel estimator. With this in mind, we caution against using our methods in extremely high-dimensional settings, since rates of convergence will be slow.

Next we present a nearly matching lower bound, proven in Appendix A through a LeCam two-point argument that extends lower bound constructions for estimating Lipschitz functions at a point to the off-policy evaluation setting (Wainwright, 2019).

Theorem 5.2. *Let ψ_2^- be as in Theorem 4.5, suppose that the covariates X_i takes values in $[-1, 1]^d$, and that the support of the policy π_b is $[-1/2, 1/2]^d$. Then, if $n \geq 2^{1-d}L^2$, the minimax risk $\inf_{\hat{\psi}_2^-} \sup_{P \in \mathcal{P}_L^{\mathrm{Lip}}} \mathbb{E}_P\left[(\hat{\psi}_2^- - \psi_2^-)^2\right]$ is at least*

$$\frac{1}{16}\mathbb{E}[\pi_e(X_i)\mathbf{1}\{\pi_b(X_i) = 0\}]^2(2n)^{-2/(d+2)}(4L)^{-2d/(d+2)}.$$

The assumption on the support of π_b simplifies the lower bound construction; we expect similar rates in other geometries. The lower bound of Theorem 5.2 shows that any estimator must have an MSE that is at least of order $n^{-2/(d+2)}$, and so the rate achieved by $\hat{\psi}_2^-$ is optimal up to log-factors. To see why we obtain this rate, note that ψ_2^- depends on a supremum of μ over the overlap region. Thus, we need to be able to estimate the value of μ at a particular point, namely the point at which the supremum is attained. The minimax lower bound for estimating an L-Lipschitz function in d dimensions at a point is $n^{-2/(d+2)}$, and so we inherit this rate for this estimation of ψ_2^-.

Table 1. Coverage percentage (as in Theorem 4.1 with $\epsilon = 0.01$) of partial identification intervals for the value of π_e on the Yeast dataset at a range of samples sizes n and smoothness parameters L. In parentheses are the percent rates at which Assumption 4.2 is satisfied. Results are averaged over 10,000 replications. We see that the Lipschitz assumption with $L = 1, 2$ likely does not hold, since Assumption 4.2 is not satisfied in larger sample sizes, while the Lipschitz assumption for larger values of L seems to hold, since Assumption 4.2 is satisfied and coverages approach the desired 100% rate.

L	$n = 1000$	$n = 2000$	$n = 3000$	$n = 4000$	$n = 5000$	$n = 10000$
1	40.8 (78.5)	0.03 (2.0)	0.0 (0)	0.0 (0)	0 (0)	0 (0)
2	59.7 (100)	72.5 (100)	78.9 (100)	82.7 (100)	85.9 (99.5)	9.5 (10.6)
3	64.6 (100)	77.2 (100)	82.8 (100)	86.9 (100)	89.7 (100)	96.9 (100)
4	68.2 (100)	80.2 (100)	85.3 (100)	89.1 (100)	91.9 (100)	98.0 (100)
5	70.6 (100)	82.0 (100)	86.8 (100)	90.7 (100)	93.1 (100)	98.4 (100)
∞	75.0 (100)	86.1 (100)	90.7 (100)	93.5 (100)	95.6 (100)	99.2 (100)

Based on this intuition, we expect that similar lower bounds will continue to hold for any non-parametric assumption that requires estimation of μ in the overlap region, such as monotonicity assumptions or higher-order smoothness assumptions. In contrast, we do not need to estimate μ when making the boundedness assumption of $\mathcal{P}_{\ell,u}^{\mathrm{bdd}}$, and so the resulting bounds will converge at a faster n^{-1} rate. As a consequence, the inferential methods of (Imbens and Manski, 2004) under the boundedness assumption do not extend to more complex smoothness assumptions; developing inferential methods under smoothness assumptions is therefore an interesting direction for future work.

6. Experiments

We now demonstrate our methods in two semi-synthetic settings. In the first setting, we study the coverage guarantees of Theorem 4.1 and compare the width of our intervals to the width of the intervals of Ben-Michael et al. (2021); in the second, we demonstrate the utility of our methods in a real-world setting.

[1]

For another example of how our methods perform on real data, we refer interested readers to Saveski et al. (2023), which takes a partial identification approach to evaluating policies in peer review management systems.

Yeast dataset: coverage analysis. Following prior work on OPE, we convert a classic multi-class classification dataset into an off-policy evaluation dataset (Dudík et al., 2011; Wang et al., 2017; Wu and Wang, 2018; Su et al., 2020; Zhan et al., 2021). Like those prior works, we use the Yeast data set from the UCI repository (Dua and Graff,

[1]Code for replicating our results is available at `https://github.com/skhan1998/lipschitz-ope`.

2017), which consists of $n = 1,484$ observations of data points $(X_i, \tilde{Y}_i)$, where $X_i \in [0,1]^8$ is a covariate vector of length $d = 8$ and $\tilde{Y}_i$ is a class label indicating one of 10 classes. We remove 6 rare classes from the data, leaving $n = 1,299$ observations distributed among 4 classes.

We sample with replacement from the observed data to generate samples of different sizes, n (so P_0 is the empirical distribution of the data). The action set $\mathcal{A}$ is the set of 4 classes in the data, and the evaluation policy, π_e, samples actions from the fitted probabilities of a logistic regression. The behavior policy, π_b, samples actions from the fitted probabilities of the same logistic regression, but with probabilities below 0.05 set to zero. This configuration leads to an overlap violation between the behavior and evaluation policies. The outcomes Y_i are 1 when the action taken matches the true class label $\tilde{Y}_i$, and 0 otherwise.

We investigate the guarantees of Theorem 4.1 and the role of the parameter L. We can generate partial identification intervals for the value of π_e under the assumption that $P_0 \in \mathcal{P}_L^{\text{Lip}} \cap \mathcal{P}_{0,1}^{\text{bdd}}$ using the methods of Section 4, where the Lipschitz assumption is made with respect to Euclidean distance. We fit $\hat{\mu}$ using a regularized logistic regression. Table 1 shows the coverage rates of these intervals, in the sense of Theorem 4.1 with $\epsilon = 0.01$, for a range of values of L. In parentheses are the fraction of times that $\hat{\mu}$ is L-Lipschitz in the overlap region, i.e., the fraction of times that Assumption 4.2 is satisfied. When Assumption 4.2 is not satisfied, the partial identification bounds are undefined, and so we never cover the off-policy value. Note that $L = \infty$ corresponds to pure Manski bounding, or equivalently to assuming only that $P_0 \in \mathcal{P}_{0,1}^{\text{bdd}}$, an assumption that is satisfied by construction in this example. We provide a plot of the results of Table 1, as well as results with $\epsilon = 0.005$, in Appendix C.

In Table 1, we see a distinction between $L = 1, 2$ and $L > 2$. For the small values of L, the problem gradually ceases to become feasible for larger values of n, and so the resulting intervals are undefined and rarely cover the off-policy value. For larger values of L, the problem is always feasible at all values of n, and the coverage increases with n, approaching the desired 100% coverage in large sample sizes. In particular, the fact that the coverage of intervals constructed assuming, for example, that $L = 4$ or $L = 5$ (an assumption that is not satisfied by construction) is close to the coverage of intervals satisfied by the $L = \infty$ boundedness assumption (which is satisfied by construction), suggests that the smoothness assumption for these larger values of L is quite plausible. As a point of reference, the $L = 5$ assumption on these covariates in the Euclidean metric implies, for example, that if two units agree in all but one covariate, and disagree by 0.2 in that covariate, their expected outcome can differ by 1, so they do not place any bounds on each other.

Table 2. Average width of the intervals obtained using the proposed method in semi-synthetic experiments on the Yeast dataset. We compare to the intervals of a baseline method proposed by Ben-Michael et al. (2021) and compute the ratio between the width of the two intervals. We set $L = 1$, average over 1000 trials, and scale the results by 10^3 for readability. We find that the intervals of baseline method are 40-50% wider than the proposed method.

	Sample size, n					
Method	500	1000	1500	2000	2500	3000
Proposed	4.37	3.09	2.59	2.15	1.80	1.68
Baseline	6.57	4.89	3.89	3.18	2.64	2.41
Ratio	1.50	1.58	1.50	1.48	1.47	1.43

Yeast dataset: interval width analysis. We now study the width of our intervals, using the intervals proposed by Ben-Michael et al. (2021) as a baseline, on the Yeast data. This baseline method constructs lower bounds on the off-policy value using a simultaneous confidence interval for $\hat{\mu}$, rather than the fitted $\hat{\mu}$ itself. We use the same dataset and construct the behavior and evaluation policies in the same way as in the previous section, but discretize the original covariates X_i according to the map $X_i \mapsto \mathbf{1}\{X_i < 0.5\}$. We discretize since, as the authors descibe, constructing the simultaneous confidence interval needed by the baseline method is challenging in the case of continuous covariates.

Table 2 shows the results for $L = 1$ and a range of sample sizes. We find that the intervals obtained using the baseline methods are, in this setting, consistently 40-50% wider than those obtained using the proposed method. This conservativeness is induced by the simultaneous confidence interval construction, which the baseline method relies on to obtain better theoretical guarantees for policy learning, but is needlessly loose for policy evaluation.

Yahoo! Front Page Today dataset. Our second experiment uses the Yahoo Webscope's featured news dataset, a standard benchmark for OPE algorithms (Yahoo!, 2011; Li et al., 2010; 2011). This dataset was collected over 10 days in May 2009, and consists of observations of user visits and actions on the front page of the Yahoo! web portal. Each observation is a tuple (X_i, A_i, Y_i), where $X_i \in [0,1]^5$, A_i is an article shown to the user in a featured position on the page, and Y_i is a binary indicator whether the user clicked. Each A_i is accompanied by a 5-dimensional covariate vector V_i. The articles A_i are chosen from a hand-curated pool of articles that is updated every hour. We restrict our focus to a single hour of data so that the articles are drawn from a fixed pool of articles $\mathcal{A}$, considering $n = 16,628$ data points. This problem thus has the form of a multi-armed bandit problem, as described in Section 3, with action space $\mathcal{A}$.

The architects of this dataset sampled $A_i \sim \text{Unif}(\mathcal{A})$, which guarantees overlap, but requires taking many sub-optimal actions. A tuned, non-uniform logging policy would clearly be preferable. With our methods, we show that it is possible to run a non-uniform policy and yet still evaluate other policies that do not satisfy overlap.

We consider the following scenario: from historical data, we know that the user covariate $X_{i,3}$ is positively correlated with the response Y_i and that the article covariate $V_{i,0}$ is positively correlated with Y_i. Based on this, we believe that we should avoid showing articles with low values of $V_{i,0}$ to users with low values of $X_{i,3}$. Thus we consider the family of policies

$$\pi^{(T)}(X_i, a) = \begin{cases} 1/|\mathcal{A}| & \text{if} \quad X_{i,3} > T, \\ \mathbf{1}\{a \in \mathcal{A}^*\}/|\mathcal{A}^*| & \text{if} \quad X_{i,3} \leq T, \end{cases}$$

where $\mathcal{A}^*$ is a subset of articles we expect to perform well and T is a cut-off that identifies users who are unlikely to click. Here, we take $\mathcal{A}^*$ to be the set of articles with above median values of $V_{i,0}$.

We suppose that we have deployed $\pi^{(0.5)}$ (and simulate this by subsampling), and are interested in exploring other values of T. The subsampled dataset contains $n = 10,086$ data points. For values of $T \geq 0.5$, the behavior policy provides full support for the evaluation policy, and no partial identification is required. For values of $T < 0.5$, there is an overlap violation, since $\pi^{(T)}$ can show users with $X_{i,3} \in (T, 0.5)$ articles $a \in \mathcal{A} \setminus \mathcal{A}^*$, an action that $\pi^{(0.5)}$ assigns zero probability.

We use the estimators of Section 4 to construct interval estimates under the assumption that $P_0 \in \mathcal{P}_L^{\text{Lip}} \cap \mathcal{P}_{0,1}^{\text{bdd}}$, where the Lipschitz assumption is made with respect to Euclidean distance. Our results are in Figure 3, which plots the interval estimators $[\hat{\psi}^-, \hat{\psi}^+]$ of the value of $\pi^{(T)}$ under Lipschitz assumptions for $T = 0.25, 0.3, 0.35, 0.4$, with $\hat{\mu}$ fit using a regularized logistic regression (which leads to smooth $\hat{\mu}$), for a range of L. Results with larger values of L and T can be found in Appendix C. As a point of reference, the $L = 1$ assumption on this data implies, for example, that if two units agree in all but one covariate, and differ by the maximum possible value of 1 in that covariate, their expected outcome can differ by the maximum possible value of 1, so they place no bounds on each other.

Also shown are the point estimates obtained by using model predictions without any partial identification, as well as an infeasible sample estimate (and accompanying confidence intervals) of the value of $\pi^{(T)}$ as estimated on data from a uniform policy. This infeasible sample estimate corresponds to the results of the experiment we would run (at considerable cost) if unwilling to rely on smoothness assumptions.

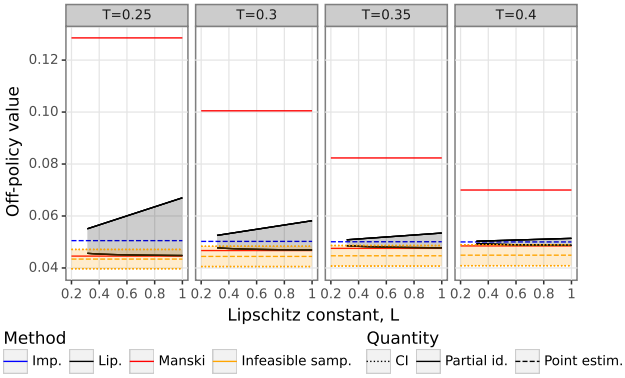

Figure 3. Partial identification bounds assuming that $P_0 \in \mathcal{P}_L^{\text{Lip}} \cap \mathcal{P}_{0,1}^{\text{bdd}}$ (black), Manski bounds assuming that $P_0 \in \mathcal{P}_{0,1}^{\text{bdd}}$ (red), and pure imputation estimates (blue) of the value of $\pi^{(T)}$ estimated using data from the behavior policy $\pi^{(0.5)}$ for a range of T and L. The point estimate and confidence intervals from an infeasible sample dataset from the uniform behavior policy are in orange. Imputation overestimates the value of $\pi^{(T)}$, and our bounds correct for this—the correction is more aggressive as the model estimate and infeasible sample estimate diverge, meaning we correctly adjust for the model's extrapolation power. There are overlap violations for 16.8%, 10.8%, 7.0%, and 4.3% of the points when $T = 0.25, 0.3, 0.35$, and 0.4 respectively.

The key takeaway from Figure 3 is that, as T increases, and the imputation estimate and infeasible sample estimate grow closer, the width of our partial identification interval decreases. This is because, as T becomes smaller, the number of units in the no-overlap region $\{X_i : T \leq X_{i,3} \leq 0.5\}$ increases, and the maximum distance between the overlap region the no-overlap region increases. Our method accounts for this, and correctly distinguishes cases where the model needs to be corrected only slightly from cases where it must be corrected substantially. Furthermore, for larger values of L, our intervals consistently overlap with the 95% confidence interval from the alternative ("costly") experiment under a uniform policy.

We also compare to the Manski partial identification regions obtained solely from the assumption that $P \in \mathcal{P}_{0,1}^{\text{bdd}}$, shown in red. One may be concerned that the intersection between our intervals and the 95% confidence interval is not large, but because this is true for the Manski intervals as well (whose non-parametric assumption holds exactly), we conclude that the small intersection is due to error in the estimation of $\hat{\psi}_1$. More importantly, we see that our intervals are much narrower than the Manski intervals for small L and converge to them as $L \to \infty$. For example, when $L = 1$, our intervals are 73.5%, 79.1%, 83.7%, and 91.5% narrower than the Manksi intervals for $T = 0.25, 0.3, 0.35$, and 0.4 respectively.

7. Conclusion

We have developed partial identification results for off-policy evaluation under smoothness assumptions, shown that our bounds have favorable asymptotics, and demonstrated their value in experiments. We emphasize that our methods are tied to a choice of covariate space and metric, and so it is crucial to evaluate the reasonability of the smoothness assumption with regard to those choices. The bounds we give are only as useful as the underlying assumptions are plausible, and thus it is essential to combine our methodological proposals with thoughtful application.

The proposed methods open many avenues for future work. One promising direction is extending our methods to assumptions made in action covariate space rather than user covariate space and to other kinds of nonparametric assumptions such as monotonicity, convexity, other kinds of smoothness, or combinations of these. In such cases, the no-interaction property may no longer hold (see Appendix D for such an example), necessitating solving a linear program numerically. Such numerical solutions could potentially still be analyzed using results on perturbations of linear programs, similar to the approach of Kallus et al. (2022).

It would also be of interest to study a setting in which there are points for which $\pi_b(X_i)$ is extremely small but non-zero. In this setting, treating such points as though they are in fact no-overlap and partially identifying their contribution to the off-policy value (rather than point identifying it using inverse-propensity based methods) may lead to tighter bounds on the off-policy value. The main challenge here is selecting a cut-off at which $\pi_b(X_i)$ should be treated as "effectively zero," possibly in a data-adaptive way, and accounting for this choice in the resulting intervals.

Acknowledgements

This work was supported in part by NSF CAREER Award #2143176. We are grateful to Kevin Guo, Tal Wagner, Nihar Shah, and Steven Jecmen for helpful discussions.

Impact Statement

This paper presents work whose goal is to advance the field of machine learning. There are many potential societal consequences of our work, none which we feel must be specifically highlighted here.

References

A. Andoni, P. Indyk, and I. Razenshteyn. Approximate nearest neighbor search in high dimensions. In *Proceedings of the International Congress of Mathematicians: Rio de Janeiro 2018*, pages 3287–3318. World Scientific, 2018.

E. Ben-Michael, D. J. Greiner, K. Imai, and Z. Jiang. Safe policy learning through extrapolation: Application to pretrial risk assessment. *arXiv preprint arXiv:2109.11679*, 2021.

J. L. Bentley. Multidimensional binary search trees used for associative searching. *Communications of the ACM*, 18 (9):509–517, 1975.

L. Bottou, J. Peters, J. Quiñonero-Candela, D. X. Charles, D. M. Chickering, E. Portugaly, D. Ray, P. Simard, and E. Snelson. Counterfactual reasoning and learning systems: The example of computational advertising. *Journal of Machine Learning Research*, 14(11), 2013.

A. Chin, D. Eckles, and J. Ugander. Evaluating stochastic seeding strategies in networks. *Management Science*, 68 (3):1714–1736, 2022.

D. Dua and C. Graff. UCI machine learning repository, 2017. URL http://archive.ics.uci.edu/ml.

M. Dudík, J. Langford, and L. Li. Doubly robust policy evaluation and learning. In *Proceedings of the 28th International Conference on International Conference on Machine Learning*, pages 1097–1104, 2011.

S. Har-Peled. *Geometric approximation algorithms*. Number 173. American Mathematical Soc., 2011.

T. Hastie, R. Tibshirani, J. H. Friedman, and J. H. Friedman. *The elements of statistical learning: data mining, inference, and prediction*, volume 2. Springer, 2009.

S. Higbee. Policy learning with new treatments. *arXiv preprint arXiv:2210.04703*, 2022.

G. W. Imbens and C. F. Manski. Confidence intervals for partially identified parameters. *Econometrica*, 72(6):1845–1857, 2004.

N. Jiang and J. Huang. Minimax value interval for off-policy evaluation and policy optimization. *Advances in Neural Information Processing Systems*, 33:2747–2758, 2020.

N. Kallus, X. Mao, and A. Zhou. Assessing algorithmic fairness with unobserved protected class using data combination. *Management Science*, 68(3):1959–1981, 2022.

L. LeCam. Convergence of estimates under dimensionality restrictions. *The Annals of Statistics*, pages 38–53, 1973.

Y. T. Lee and A. Sidford. Efficient inverse maintenance and faster algorithms for linear programming. In *2015 IEEE 56th Annual Symposium on Foundations of Computer Science*, pages 230–249. IEEE, 2015.

L. Li, W. Chu, J. Langford, and R. E. Schapire. A contextual-bandit approach to personalized news article recommendation. In *Proceedings of the 19th international conference on World wide web*, pages 661–670, 2010.

L. Li, W. Chu, J. Langford, and X. Wang. Unbiased offline evaluation of contextual-bandit-based news article recommendation algorithms. In *Proceedings of the fourth ACM international conference on Web search and data mining*, pages 297–306, 2011.

P. Liao, P. Klasnja, and S. Murphy. Off-policy estimation of long-term average outcomes with applications to mobile health. *Journal of the American Statistical Association*, 116(533):382–391, 2021.

C. F. Manski. Nonparametric bounds on treatment effects. *The American Economic Review*, 80(2):319–323, 1990.

W. Mou, P. Ding, M. J. Wainwright, and P. L. Bartlett. Kernel-based off-policy estimation without overlap: Instance optimality beyond semiparametric efficiency. *arXiv preprint arXiv:2301.06240*, 2023.

S. M. Omohundro. *Five balltree construction algorithms*. International Computer Science Institute Berkeley, 1989.

F. P. Preparata and M. I. Shamos. *Computational geometry: an introduction*. Springer Science & Business Media, 2012.

N. Sachdeva, Y. Su, and T. Joachims. Off-policy bandits with deficient support. In *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 965–975, 2020.

M. Saveski, S. Jecmen, N. B. Shah, and J. Ugander. Counterfactual evaluation of peer-review assignment policies. *Advances in Neural Information Processing Systems*, 2023.

C. J. Stone. Optimal global rates of convergence for nonparametric regression. *The annals of statistics*, pages 1040–1053, 1982.

J. Stoye. More on confidence intervals for partially identified parameters. *Econometrica*, 77(4):1299–1315, 2009.

Y. Su, M. Dimakopoulou, A. Krishnamurthy, and M. Dudík. Doubly robust off-policy evaluation with shrinkage. In *International Conference on Machine Learning*, pages 9167–9176. PMLR, 2020.

A. Swaminathan and T. Joachims. Batch learning from logged bandit feedback through counterfactual risk minimization. *The Journal of Machine Learning Research*, 16(1):1731–1755, 2015.

A. Swaminathan, A. Krishnamurthy, A. Agarwal, M. Dudik, J. Langford, D. Jose, and I. Zitouni. Off-policy evaluation for slate recommendation. *Advances in Neural Information Processing Systems*, 30, 2017.

P. Thomas and E. Brunskill. Data-efficient off-policy policy evaluation for reinforcement learning. In *International Conference on Machine Learning*, pages 2139–2148. PMLR, 2016.

M. J. Wainwright. *High-dimensional statistics: A non-asymptotic viewpoint*, volume 48. Cambridge university press, 2019.

Y.-X. Wang, A. Agarwal, and M. Dudık. Optimal and adaptive off-policy evaluation in contextual bandits. In *International Conference on Machine Learning*, pages 3589–3597. PMLR, 2017.

H. Wu and M. Wang. Variance regularized counterfactual risk minimization via variational divergence minimization. In *International Conference on Machine Learning*, pages 5353–5362. PMLR, 2018.

T. Xie, C.-A. Cheng, N. Jiang, P. Mineiro, and A. Agarwal. Bellman-consistent pessimism for offline reinforcement learning. *Advances in neural information processing systems*, 34:6683–6694, 2021.

Yahoo! Webscope dataset R6A: ydata-frontpage-todaymodule-clicks-v1, 2011. URL https://webscope.sandbox.yahoo.com/.

R. Zhan, V. Hadad, D. A. Hirshberg, and S. Athey. Off-policy evaluation via adaptive weighting with data from contextual bandits. In *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*, pages 2125–2135, 2021.

Off-policy Evaluation Beyond Overlap: Sharp Partial Identification Under Smoothness

Supplemental material

A. Proofs of results

A.1. Proof of Theorem 4.1

Proof of Theorem 4.1. We begin by writing

$$\mathbb{P}(\hat{\psi}^- - \epsilon \leq \psi(P_0) \leq \hat{\psi}^+ + \epsilon) = 1 - \mathbb{P}(\psi(P_0) < \hat{\psi}^- - \epsilon \text{ or } \psi(P_0) > \hat{\psi}^+ + \epsilon), \tag{5}$$

$$\geq 1 - \mathbb{P}(\psi(P_0) \leq \hat{\psi}^- - \epsilon) - \mathbb{P}(\psi(P_0) > \hat{\psi}^+ + \epsilon), \tag{6}$$

by the union bound. We now show that $\mathbb{P}(\psi(P_0) > \hat{\psi}^+ + \epsilon) \to 0$. An analogous argument shows that $\mathbb{P}(\psi(P_0) < \hat{\psi}^- - \epsilon) \to 0$ as well, and these two facts along with (6) imply the result.

We have

$$\mathbb{P}(\psi(P_0) > \hat{\psi}^+ + \epsilon) = \mathbb{P}(\psi_1(P_0) + \psi_2(P_0) > \hat{\psi}_1 + \hat{\psi}_2^+ + \epsilon), \tag{7}$$

$$\leq \mathbb{P}(\psi_1(P_0) > \hat{\psi}_1 + \epsilon/2) + \mathbb{P}(\psi_2(P_0) > \hat{\psi}_2^+ + \epsilon/2), \tag{8}$$

$$\leq \mathbb{P}(\psi_1(P_0) > \hat{\psi}_1 + \epsilon/2) + \mathbb{P}\left(\sup_{P \in \mathcal{P}} \psi_2(P) > \hat{\psi}_2^+ + \epsilon/2\right), \tag{9}$$

$$\leq \mathbb{P}(|\psi_1(P_0) - \hat{\psi}_1| > \epsilon/2) + \mathbb{P}\left(\left|\sup_{P \in \mathcal{P}} \psi_2(P) - \hat{\psi}_2^+\right| > \epsilon/2\right). \tag{10}$$

Both terms of (10) are $o(1)$ under the given consistency assumptions, so we conclude that $\mathbb{P}(\psi(P_0) > \hat{\psi}^+ + \epsilon) = o(1)$ as desired. □

A.2. Proof of Theorem 4.5

We in fact prove a generalization of Theorem 4.5 that is based on the assumption that $P_0 \in \mathcal{P}_L^{\text{Lip}} \cap \mathcal{P}_{\ell,u}^{\text{bdd}}$. In particular, consider the optimization problem

$$\begin{aligned} \min_{t_1, \cdots, t_n} \quad & \frac{1}{n} \sum_{i=1}^n t_i \pi_e(X_i) \mathbf{1}\{\pi_b(X_i) = 0\} \\ \text{s.t.} \quad & |t_i - t_j| \leq L d(X_i, X_j), \quad 1 \leq i < j \leq n, \\ & t_i - \hat{\mu}(X_i) = 0, \quad 1 \leq i \leq n \text{ s.t. } \pi_b(X_i) > 0, \\ & \ell \leq t_i \leq u, \quad 1 \leq i \leq n \end{aligned} \tag{11}$$

Then we have the following result.

Theorem A.1. *Suppose that (11) is feasible, that $P_0 \in \mathcal{P}_L^{\text{Lip}}$, and that Assumptions 4.3 and 4.4 hold. Then:*

(a) the problem (11) has value

$$\hat{\psi}_2^- = \frac{1}{n} \sum_{i=1}^n \pi_e(X_i) \left(\ell \vee \max_{j : \pi_b(X_j) > 0} \hat{\mu}(X_j) - L d(X_i, X_j) \right) \mathbf{1}\{\pi_b(X_i) = 0\}; \tag{12}$$

(b) we have $\hat{\psi}_2^- \xrightarrow{\mathbb{P}} \psi_2^{-,\infty}$ where

$$\psi_2^{-,\infty} = \mathbb{E}\left[\pi_e(X_i) \left(\ell \vee \sup_{x : \pi_b(x) > 0} \mu_{P_0}(x) - L d(X_i, x) \right) \mathbf{1}\{\pi_b(X_i) = 0\} \right]; \tag{13}$$

(c) the bound $\psi_2^{-,\infty}$ is sharp in the sense that $\inf_{P \in \mathcal{P}_L^{\text{Lip}} \cap \mathcal{P}_M^{\text{bdd}}} \psi_2(P) = \psi_2^{-,\infty}$.

Theorem 4.5 from the main text follows from sending $\ell \to -\infty$ in Theorem A.1. We now prove each part of the theorem in turn. In our proofs, we assume for the sake of convenience that the supremum in (13) is attained. This will be the case if, for instance, $\{x : \pi_b(x) > 0\}$ is compact. If the supremum is not attained, slight modifications of our proofs can be used to obtain the result.

Proof of Theorem A.1(a). To show the result, we must do two things: show that the objective of (3) is bounded below by $\hat{\psi}_2^-$, and show that this bound is attained.

For the bound, observe that, for each t_i, we have

$$\begin{aligned} t_i &\geq t_j - Ld(X_i, X_j) \text{ for } 1 \leq j \leq n, && (14)\\ &\geq \max_{j:\pi_b(X_j)>0} t_j - Ld(X_i, X_j), && (15)\\ &\geq \max_{j:\pi_b(X_j)>0} t_j - Ld(X_i, X_j), && (16)\\ &= \max_{j:\pi_b(X_j)>0} \hat{\mu}(X_j) - Ld(X_i, X_j), && (17) \end{aligned}$$

where the first bound is the Lipschitz constraint and the last equality is the equality constraint. We also have $t_i \geq \ell$ for all i, and thus we must have

$$\sum_{i=1}^n t_i \pi_e(X_i) \mathbf{1}\{\pi_b(X_i) = 0\} \geq \sum_{i=1}^n \pi_e(X_i) \left(\ell \vee \max_{j:\pi_b(X_j)>0} \hat{\mu}(X_j) - Ld(X_i, X_j) \right) \mathbf{1}\{\pi_b(X_i) = 0\} = \hat{\psi}_2^-, \tag{18}$$

as desired.

Next we will show that there exist a set of feasible t_i for which the objective of (3) is equal to $\hat{\psi}_2^-$. The construction is

$$t_i^* = \ell \vee \max_{j:\pi_b(X_j)>0} \hat{\mu}(X_j) - Ld(X_i, X_j). \tag{19}$$

This construction clearly has objective value $\hat{\psi}_2^-$, so we need only check that it is feasible.

For the equality constraint, note that if $\pi_b(X_i) > 0$, then the maximum in (19) includes $j = i$, and thus is at least $\hat{\mu}(X_i)$. But for any $j \neq i$, we have $|\hat{\mu}(X_j) - \hat{\mu}(X_i)| \leq Ld(X_i, X_j)$ by the feasibility of (3), and so $\hat{\mu}(X_i) \geq \hat{\mu}(X_j) - Ld(X_i, X_j)$. Since $\hat{\mu}$ is range-bounded, we also have $\hat{\mu}(X_i) \geq \ell$, and so

$$\ell \vee \max_{j:\pi_b(X_j)>0} \hat{\mu}(X_j) - Ld(X_i, X_j) = \hat{\mu}(X_i), \tag{20}$$

for i in the overlap region, and the equality constraint is satisfied.

Next we check the Lipschitz constraint. To do this, we define

$$f(i) = \operatorname*{argmin}_{j:\pi_b(X_j)>0} \hat{\mu}(X_j) - Ld(X_i, X_j) \tag{21}$$

so that $t_i^* = \ell \vee (\hat{\mu}(X_{f(i)}) - Ld(X_i, X_{f(i)}))$. We consider any two t_i^*, t_j^*, and distinguish three cases.

First, if $\hat{\mu}(X_{f(i)}) - Ld(X_i, X_{f(i)}) < \ell$ and $\hat{\mu}(X_{f(j)}) - Ld(X_j, X_{f(j)}) < \ell$, then $|t_i^* - t_j^*| = 0$ and the Lipschitz condition is satisfied.

Second, if $\hat{\mu}(X_{f(i)}) - Ld(X_i, X_{f(i)}) > \ell$ and $\hat{\mu}(X_{f(j)}) - Ld(X_j, X_{f(j)}) > \ell$, we may assume without the loss of generality that $t_i^* > t_j^*$, and compute

$$\begin{aligned} |t_i^* - t_j^*| &= t_i^* - t_j^*, && (22)\\ &= \left(\hat{\mu}(X_{f(i)}) - Ld(X_i, X_{f(i)})\right) - \left(\hat{\mu}(X_{f(j)}) - Ld(X_j, X_{f(j)})\right), && (23)\\ &\leq \left(\hat{\mu}(X_{f(i)}) - Ld(X_i, X_{f(i)})\right) - \left(\hat{\mu}(X_{f(i)}) - Ld(X_j, X_{f(i)})\right), && (24)\\ &= L\left(d(X_j, X_{f(i)}) - d(X_i, X_{f(i)})\right), && (25)\\ &\leq Ld(X_i, X_j), && (26) \end{aligned}$$

where the first inequality uses the maximality (by definition) of $f(j)$, and the second inequality is the triangle inequality $d(X_j, X_{f(i)}) \leq d(X_i, X_{f(i)}) + d(X_i, X_j)$. Thus the Lipschitz constraint is satisfied in this case.

Finally, if $\hat{\mu}(X_{f(i)}) - Ld(X_i, X_{f(i)}) > \ell$ and $\hat{\mu}(X_{f(j)}) - Ld(X_j, X_{f(j)}) < \ell$, then

$$|t_i^* - t_j^*| = \hat{\mu}(X_{f(i)}) - Ld(X_i, X_{f(i)}) - (\ell), \tag{27}$$

$$\leq \hat{\mu}(X_{f(i)}) - Ld(X_i, X_{f(i)}) - (\hat{\mu}(X_{f(j)}) - Ld(X_j, X_{f(j)})) \tag{28}$$

and (28) is bounded by $Ld(X_i, X_j)$ by the arguments of (26).

Thus we see that the t_i^* defined in (19) are feasible, and conclude that the value of (3) is $\hat{\psi}_2^-$. □

Before proceeding to the proof of Theorem A.1(b), we present two helpful lemmas. The first of these controls the difference between the suprema of interest in terms of the difference in conditional mean functions, and will essentially be used to replace the estimated $\hat{\mu}$ in $\hat{\psi}_2^-$ with the true μ.

Lemma A.2. *For any conditional mean functions μ_1, μ_2, we have*

$$\left| \max_{j:\pi_b(X_j)>0} \mu_1(X_j) - Ld(X_i, X_j) - \max_{j:\pi_b(X_j)>0} \mu_2(X_j) - Ld(X_i, x) \right| \leq \sup_{x:\pi_b(x)>0} |\mu_1(x) - \mu_2(x)|. \tag{29}$$

Proof of Lemma A.2. We have

$$\max_{j:\pi_b(X_j)>0} \mu_1(X_j) - Ld(X_i, X_j) \leq \max_{j:\pi_b(X_j)>0} \mu_1(X_j) - \mu_2(X_j) + \max_{j:\pi_b(X_j)>0} \mu_2(X_j) - Ld(X_i, X_j), \tag{30}$$

$$\leq \sup_{x:\pi_b(x)>0} |\mu_1(x) - \mu_2(x)| + \max_{j:\pi_b(X_j)>0} \mu_2(X_j) - Ld(X_i, X_j), \tag{31}$$

$$\tag{32}$$

The same bounds hold with the roles of μ_1 and μ_2 reversed, completing the proof. □

The next two lemmas will allow us to replace the maximum in $\hat{\psi}_2^-$ by a supremum by controlling the difference between the maximum and supremum.

Lemma A.3. *Let x^* be the point at which $\sup_{x:\pi_b(x)>0} \mu_{P_0}(x) - Ld(X_i, x)$ is attained and let*

$$j^* = \operatorname*{argmin}_{j:\pi_b(X_j)>0} d(X_j, x^*) \tag{33}$$

be the observed point that is closest to x^*. Then

$$\left| \max_{j:\pi_b(X_j)>0} \mu_{P_0}(X_j) - Ld(X_i, X_j) - \sup_{x:\pi_b(x)>0} \mu_{P_0}(x) - Ld(X_i, x) \right| \leq 2Ld(x^*, X_{j^*}). \tag{34}$$

Proof. We have

$$\mu_{P_0}(X_{j^*}) - Ld(X_i, X_{j^*}) \geq \mu_{P_0}(x^*) - Ld(X_{j^*}, x^*) - Ld(X_i, X_{j^*}), \tag{35}$$

$$\geq \mu_{P_0}(x^*) - Ld(X_i, x^*) - 2Ld(x^*, X_{j^*}), \tag{36}$$

$$= \sup_{x:\pi_b(x)>0} \mu_{P_0}(x) - Ld(X_i, x) - 2Ld(x^*, X_{j^*}), \tag{37}$$

where the first inequality uses the fact that μ_{P_0} is L-Lipschitz and the second is the triangle inequality. Since the supremum is greater than the maximium, the result follows. □

Lemma A.4. *Fix a point x such that $\pi_b(x) > 0$ and let*

$$j^* = \operatorname*{argmin}_{j:\pi_b(X_j)>0} d(X_j, x) \tag{38}$$

be the index of the closest observation to x in the overlap region. Then, under either of Assumption 4.4(a) or Assumption 4.4(b), we have $\mathbb{P}(d(x, X_{j^*}) > \epsilon) \xrightarrow{n\to\infty} 0$ for any $\epsilon > 0$.

Proof. We analyze $\mathbb{P}(d(x, X_{j^*}) > \epsilon)$ under each of the two possibilities in Assumption 4.4. If Assumption 4.4(a) holds and there is an atom at x so that $\mathbb{P}(X_i = x) = p$ for some $p > 0$, we have

$$\mathbb{P}(d(x, X_{j^*}) > \epsilon) \leq \mathbb{P}(X_i \neq x \text{ for } 1 \leq i \leq n), \tag{39}$$
$$= (1-p)^n, \tag{40}$$

which goes to 0 as $n \to \infty$. Next, if Assumption 4.4(b) holds, let δ be such that $\mathbb{P}(d(x, X_{j^*}) \leq \epsilon) > \delta$. Then

$$\mathbb{P}(d(x, X_{j^*}) > \epsilon) = \mathbb{P}(d(x, X_i) > \epsilon \text{ for } 1 \leq i \leq n), \tag{41}$$
$$\leq (1-\delta)^n \tag{42}$$

which again goes to 0 as $n \to \infty$. □

With these lemmas in hand, we begin the main proof.

Proof of Theorem A.1(b). We begin by defining

$$\hat{\psi}_2^{-,\text{oracle}} = \frac{1}{n}\sum_{i=1}^{n} \pi_e(X_i)\left(\ell \vee \sup_{x:\pi_b(x)>0} \mu_{P_0}(x) - Ld(X_i, x)\right)\mathbf{1}\{\pi_b(X_i) = 0\}. \tag{43}$$

The main idea is to show that

$$|\hat{\psi}_2^{-} - \hat{\psi}_2^{-,\text{oracle}}| = o_P(1), \tag{44}$$

and then since $\hat{\psi}_2^{-,\text{oracle}}$ is a sum of i.i.d. terms, the desired result follows by the law of large numbers.

To show (44), we show that for any fixed i,

$$\left|\ell \vee \max_{j:\pi_b(X_j)>0} \hat{\mu}(X_j) - Ld(X_i, X_j) - \left(\ell \vee \sup_{x:\pi_b(x)>0} \mu_{P_0}(x) - Ld(X_i, x)\right)\right| = o_P(1). \tag{45}$$

Indeed, we have by the triangle inequality that

$$\left|\ell \vee \max_{j:\pi_b(X_j)>0} \hat{\mu}(X_j) - Ld(X_i, X_j) - \left(\ell \vee \sup_{x:\pi_b(x)>0} \mu_{P_0}(x) - Ld(X_i, x)\right)\right| \tag{46}$$
$$\leq \left|\ell \vee \max_{j:\pi_b(X_j)>0} \hat{\mu}(X_j) - Ld(X_i, X_j) - \left(\ell \vee \max_{j:\pi_b(X_j)>0} \mu_{P_0}(X_j) - Ld(X_i, X_j)\right)\right| \tag{47}$$
$$+ \left|\ell \vee \max_{j:\pi_b(X_j)>0} \mu_{P_0}(X_j) - Ld(X_i, X_j) - \left(\ell \vee \sup_{x:\pi_b(x)>0} \mu_{P_0}(x) - Ld(X_i, x)\right)\right|, \tag{48}$$

and now we analyze (47) and (48) separately. In both cases, we will show that if one of the terms is smaller than ℓ, the other must be as well with high probability, and then work on the event that they are both smaller than ℓ.

For (47), we first consider the event

$$E = \left\{\max_j \hat{\mu}(X_j) - Ld(X_i, X_j) > \ell, \max_j \mu_{P_0}(X_j) - Ld(X_i, X_j) < \ell\right\}. \tag{49}$$

On the event E, there must exist some j and $\epsilon > 0$ such that $\hat{\mu}(X_j) - Ld(X_i, X_j) > \ell + \epsilon$. Then,

$$\mathbb{P}(E) \leq \mathbb{P}(\hat{\mu}(X_j) - Ld(X_i, X_j) > \ell + \epsilon, \mu_{P_0}(X_j) - Ld(X_i, X_j) < \ell), \tag{50}$$
$$\leq \mathbb{P}\left(\sup_{x:\pi_b(x)>0} |\hat{\mu}(x) - \mu_{P_0}(x)| > \epsilon\right), \tag{51}$$
$$= o(1), \tag{52}$$

by Assumption 4.3.

Since $\mathbb{P}(E) = o(1)$, it is sufficient to work on the event E^c. On this event, there are two possibilities. Either

$$\mu_{P_0}(X_j) - Ld(X_i, X_j) < \ell \quad \text{and} \quad \sup_{x:\pi_b(x)>0} \mu_{P_0}(x) - Ld(X_i, x) < \ell, \tag{53}$$

or

$$\mu_{P_0}(X_j) - Ld(X_i, X_j) > \ell \quad \text{and} \quad \sup_{x:\pi_b(x)>0} \mu_{P_0}(x) - Ld(X_i, x) > \ell. \tag{54}$$

If (53) holds, then (47) is 0. On the other hand, if (54) holds, then (47) is $o_P(1)$ by Lemma A.2 and Assumption 4.3. Thus we conclude that (47) is $o_P(1)$ on the event E^c.

For (48), we distinguish cases. For the first case, if

$$\sup_{x:\pi_b(x)>0} \mu_{P_0}(x) - Ld(X_i, x) \leq \ell, \tag{55}$$

then we must have

$$\max_{j:\pi_b(X_j)>0} \mu_{P_0}(X_j) - Ld(X_i, X_j) \leq \ell, \tag{56}$$

as well, and so (48) is 0.

Thus it suffices to consider the case where

$$\sup_{x:\pi_b(x)>0} \mu_{P_0}(x) - Ld(X_i, x) > \ell. \tag{57}$$

In this case, suppose that the supremum is attained at x^*, and that $\mu_{P_0}(x^*) - Ld(X_i, x^*) - \ell = \epsilon$ for some $\epsilon > 0$, and let

$$j^* = \operatorname*{argmin}_{j:\pi_b(X_j)>0} d(X_j, x^*) \tag{58}$$

be the index of the observed data point that is closest to x^*.

Then, consider the event

$$E = \left\{ \max_{j:\pi_b(X_j)>0} \mu_{P_0}(X_j) - Ld(X_i, X_j) < \ell \right\}. \tag{59}$$

Since we are working in the case where (57) holds, by Lemma A.3, if the event E occurs as well, we must have $2Ld(x^*, X_{j^*}) > \epsilon$. Thus $\mathbb{P}(E) \leq \mathbb{P}(2Ld(x^*, X_{j^*}) > \epsilon)$, and $\mathbb{P}(2Ld(x^*, X_{j^*}) > \epsilon) = o(1)$ by Lemma A.4, so we see that $\mathbb{P}(E) = o(1)$.

Thus it suffices to work on the event E^c. On E^c, we have

$$\max_{j:\pi_b(X_j)>0} \mu_{P_0}(X_j) - Ld(X_i, X_j) \geq \sup_{x:\pi_b(x)>0} \mu_{P_0}(x) - Ld(X_i, x) - 2Ld(x^*, X_{j^*}), \tag{60}$$

by Lemma A.3. Applying Lemma A.4 again, we see that (48) is $o_P(1)$.

Since (47) and (48) are both $o_P(1)$, we conclude that (45) holds, implying (44) and finishing the proof. □

Proof of Theorem A.1(c). The proof of this result is essentially a continuous version of the proof of Theorem A.1(a). We would like to show that

$$\inf_{P \in \mathcal{P}_L^{\text{Lip}}} \mathbb{E}_P \left[Y_i \pi_e(X_i) \mathbf{1}\{\pi_b(X_i) = 0\} \right] = \mathbb{E}_{P_0} \left[\pi_e(X_i) \left(\ell \vee \sup_{x:\pi_b(x)>0} \mu_{P_0}(x) - Ld(X_i, x) \right) \mathbf{1}\{\pi_b(X_i) = 0\} \right]. \tag{61}$$

To do this, we must first show that each for each $P \in \mathcal{P}_L^{\text{Lip}} \cap \mathcal{P}_M^{\text{bdd}}$, $\psi_2(P)$ is greater than $\psi_2^{-,\infty}$, and then show that $\psi_2^{-,\infty}$ is attained.

For the lower bound, note that for any X_i and x such that $\pi_b(X_i) = 0$ and $\pi_b(x) > 0$, we must have $\mu_P(X_i) \geq \mu_P(x) - Ld(X_i, x)$ since μ_P is L-Lipschitz. Furthermore, since P is consistent with P_0 for x such that $\pi_b(x) > 0$, we in fact have $\mu_P(X_i) \geq \mu_{P_0}(x) - Ld(X_i, x)$ for all such x. Lastly, since μ_P is bounded, we also have $\mu_P(X_i) \geq \ell$. Thus,

$$\mathbb{E}_P[Y_i \pi_e(X_i) \mathbf{1}\{\pi_b(X_i) = 0\}] = \mathbb{E}_P[\pi_e(X_i)\mu_P(X_i)\mathbf{1}\{\pi_b(X_i) = 0\}], \tag{62}$$

$$\geq \mathbb{E}_P\left[\pi_e(X_i)\left(\ell \vee \sup_{x:\pi_b(x)>0} \mu_{P_0}(x) - Ld(X_i, x)\right)\mathbf{1}\{\pi_b(X_i) = 0\}\right], \tag{63}$$

$$= \mathbb{E}_{P_0}\left[\pi_e(X_i)\left(\ell \vee \sup_{x:\pi_b(x)>0} \mu_{P_0}(x) - Ld(X_i, x)\right)\mathbf{1}\{\pi_b(X_i) = 0\}\right], \tag{64}$$

where the first equality is the tower rule, the second inquality follows from the arguments of the preceeding paragraph, and the third equality uses the fact that P is consistent with P_0. Since (64) is exactly $\psi_2^{-,\infty}$, this shows the lower bound.

To show that this bound is attained, let P^* be a distribution that has the same marginal X_i distribution as P_0, the same conditional distribution of $Y_i \mid X_i = x$ for x such that $\pi_b(x) > 0$, and has

$$\mu_{P^*}(x) = \ell \vee \sup_{x':\pi_b(x')>0} \mu_{P_0}(x') - Ld(x, x'), \tag{65}$$

for x such that $\pi_b(x) = 0$. Furthermore, we assume that, for x such that $\pi_b(x) = 0$ the distribution $Y_i \mid X_i = x$ is a point mass at $\mu_{P^*}(x)$. This ensures that $P^* \in \mathcal{P}^M_{\text{bdd}}$. This P^* clearly attains the bound of $\psi_2^{-,\infty}$ so it is sufficient to check that it is consistent with P_0 and L-Lipschitz to show that $P^* \in \mathcal{P}^{\text{Lip}}_L$.

The distribution P^* is consistent with P_0 by construction, but we must still check that μ_{P^*} defined in (65) agrees with μ_{P_0} for x in the overlap region. To verify this, note that for any x such that $\pi_b(x) > 0$, the supremum in (65) is attained at $x' = x$, since $\mu_{P_0}(x) - Ld(x, x) = \mu_{P_0}(x)$ and $\mu_{P_0}(x) \geq \mu_{P_0}(x') - Ld(x, x')$ for any other x' since μ_{P_0} is L-Lipschitz. So the supremum is in fact equal to $\mu_{P_0}(x)$, and $\ell \vee \mu_{P_0}(x) = \mu_{P_0}(x)$, which is consistent with P_0.

To check that μ_{P^*} is L-Lipschitz, let $f(x)$ be the value attaining the supremum in (65), so that $\mu_{P^*}(x) = \ell \vee (\mu_{P_0}(f(x)) - Ld(x, f(x)))$. Then, consider any pair of points x_1, x_2. We distinguish three cases.

First, if $\mu_{P_0}(f(x_1)) - Ld(x_1, f(x_1)) < \ell$ and $\mu_{P_0}(f(x_2)) - Ld(x_2, f(x_2)) < \ell$, we have $\mu_{P^*}(x_1) - \mu_{P^*}(x_2) = 0$ and the Lipschitz condition is satisfied.

Second, if $\mu_{P_0}(f(x_1)) - Ld(x_1, f(x_1)) > \ell$ and $\mu_{P_0}(f(x_2)) - Ld(x_2, f(x_2)) > \ell$, assume without the loss of generality that $\mu_{P^*}(x_1) > \mu_{P^*}(x_2)$. Then we have

$$|\mu_{P^*}(x_1) - \mu_{P^*}(x_2)| = \mu_{P^*}(x_1) - \mu_{P^*}(x_2), \tag{66}$$

$$= (\mu_{P_0}(f(x_1)) - Ld(x_1, f(x_1))) - (\mu_{P_0}(f(x_2) - Ld(x_2, f(x_2))), \tag{67}$$

$$\leq (\mu_{P_0}(f(x_1)) - Ld(x_1, f(x_1))) - (\mu_{P_0}(f(x_1) - Ld(x_2, f(x_1))), \tag{68}$$

$$= L\left(d(x_1, f(x_1)) - d(x_2, f(x_1))\right), \tag{69}$$

$$\leq Ld(x_1, x_2), \tag{70}$$

where the first inequality uses the fact that $f(x_2)$ attains the supremum and the second uses the triangle inequality.

Finally, if $\mu_{P_0}(f(x_1)) - Ld(x_1, f(x_1)) > \ell$ and $\mu_{P_0}(f(x_2)) - Ld(x_2, f(x_2)) < \ell$, we have

$$|\mu_{P^*}(x_1) - \mu_{P^*}(x_2)| = (\mu_{P_0}(f(x_1)) - Ld(x_1, f(x_1))) - (\ell), \tag{71}$$

$$\leq (\mu_{P_0}(f(x_1)) - Ld(x_1, f(x_1))) - (\mu_{P_0}(f(x_2) - Ld(x_2, f(x_2))), \tag{72}$$

$$\leq Ld(x_1, x_2), \tag{73}$$

by the same arguments as above.

This shows that μ_{P^*} is L-Lipschitz, and so $P^* \in \mathcal{P}^{\text{Lip}}_L$. Thus the bound of $\psi_2^{-,\infty}$ is attained, completing the proof. □

A.3. Proof of Theorem 3

Proof. We follow the approach of the proof of Theorem A.1(b), and begin by defining

$$\hat{\psi}_2^{-,\text{oracle}} = \frac{1}{n}\sum_{i=1}^{n} \pi_e(X_i)\left(\sup_{x:\pi_b(x)>0} \mu_{P_0}(x) - Ld(X_i, x)\right)\mathbf{1}\{\pi_b(X_i) = 0\}. \tag{74}$$

We then decompose

$$\mathbb{E}[(\hat{\psi}_2^- - \psi_2^{-,\infty})^2] = \mathbb{E}[(\hat{\psi}_2^- - \hat{\psi}_2^{-,\text{oracle}} + \hat{\psi}_2^{-,\text{oracle}} - \psi_2^{-,\infty})^2], \tag{75}$$

$$\leq 2\mathbb{E}[(\hat{\psi}_2^- - \hat{\psi}_2^{-,\text{oracle}})^2] + 2\mathbb{E}[(\hat{\psi}_2^{-,\text{oracle}} - \hat{\psi}_2^{-,\infty})^2]. \tag{76}$$

The second term of (76) is the variance of an i.i.d. sum, and is thus equal to C/n for some constant C. The remainder of the proof focuses on the first term of (76).

The first term of (76) is

$$\mathbb{E}\left[\left(\frac{1}{n}\sum_{i=1}^{n}\pi_e(X_i)\left(\max_{j>\pi_b(X_j)>0}\hat{\mu}(X_j) - Ld(X_i, X_j) - \sup_{x:\pi_b(x)>0}\mu_{P_0}(x) - Ld(X_i, x)\right)\mathbf{1}\{\pi_b(X_i) = 0\}\right)^2\right], \tag{77}$$

$$\leq \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left[\left(\max_{j>\pi_b(X_j)>0}\hat{\mu}(X_j) - Ld(X_i, X_j) - \sup_{x:\pi_b(x)>0}\mu_{P_0}(x) - Ld(X_i, x)\right)^2\right], \tag{78}$$

by the Cauchy-Schwarz inequality and the fact that $0 \leq \pi_e, \mathbf{1}\{\pi_b(X_i) = 0\} \leq 1$. A single term of (78) is

$$\leq 2\mathbb{E}\left[\left(\max_{j:\pi_b(X_j)>0}\hat{\mu}(X_j) - Ld(X_i, X_j) - \max_{j:\pi_b(X_j)>0}\mu_{P_0}(X_j) - Ld(X_i, X_j)\right)^2\right] \tag{79}$$

$$+ 2\mathbb{E}\left[\left(\max_{j:\pi_b(X_j)>0}\mu_{P_0}(X_j) - Ld(X_i, X_j) - \sup_{x:\pi_b(x)>0}\mu_{P_0}(x) - Ld(X_i, x)\right)^2\right], \tag{80}$$

$$\leq 2\mathbb{E}[\|(\hat{\mu}(x) - \mu(x))\mathbf{1}\{\pi_b(x) > 0\}\|_\infty^2] + 2\mathbb{E}[L^2 d(x^*, X_{j^*})^2], \tag{81}$$

$$\leq 2\mathbb{E}[\|(\hat{\mu}(x) - \mu(x))\mathbf{1}\{\pi_b(x) > 0\}\|_\infty^2] + 4L^2(c_d bn)^{-2/d}. \tag{82}$$

where the second inequality uses Lemmas A.2 and A.3 (here x^* and X_{j^*} are as defined in Lemma A.3), and the third inequality uses Lemma A.6 (proven in Section A.4).

□

Proof. Our proof relies on LeCam's two point method (LeCam, 1973; Wainwright, 2019), which states that if we can find distributions P_1 and P_2 in $\mathcal{P}_L^{\text{Lip}}$ with $|\psi_2^{-,\infty}(P_1) - \psi_2^{-,\infty}(P_2)| \geq 2\delta$, then

$$\inf_{\hat{\psi}_2^-}\sup_{P\in\mathcal{P}_L^{\text{Lip}}}\mathbb{E}_P\left[(\hat{\psi} - \psi_2^{-,\infty}(P))^2\right] \geq \frac{\delta^2}{2}\left(1 - \|P_1^n - P_2^n\|_{\text{TV}}\right), \tag{83}$$

where by a slight abuse of notation we write P_1^n for the joint distribution of $(X_1, A_1, A_1Y_1), \cdots, (X_n, A_n, A_nY_n)$ when (X_i, Y_i) are drawn from P_1, and similarly for P_2. We construct P_1 and P_2 as follows:

(i) under P_1, the marginal distribution of X_i is uniform on $[-1, 1]^d$, and $Y_i \mid X_i = x$ is $N(\mu_1(x), 1)$ where $\mu_1(x) = 0$ identically

(ii) under P_2, the marginal distribution of X_i is uniform on $[-1, 1]^d$, and the distribution of $Y_i \mid X_i = x$ is $N(\mu_2(x), 1)$ where

$$\mu_2(x) = \begin{cases} 0 & \text{if} \quad x \in [-1/2 + L\epsilon, 1/2 - L\epsilon]^d, \\ \tilde{\mu}_2(x) & \text{if} \quad x \in [-1/2 - L\epsilon, 1/2 + L\epsilon]^d \setminus [-1/2 + L\epsilon, 1/2 - L\epsilon]^d, \\ 0 & \text{if} \quad [-1, 1]^d \setminus [-1/2 - L\epsilon, 1/2 + L\epsilon]^d, \end{cases} \tag{84}$$

where $\epsilon < 1/(2L)$ is arbitrary and $\tilde{\mu}_2(x)$ is a function that is equal to ϵ for x on the boundary of the set $[-1/2, 1/2]^d$, and is equal to 0 on the boundaries of $[-1/2-\epsilon, 1/2+\epsilon]^d$ and $[-1/2+\epsilon, 1/2-\epsilon]^d$, and linearly interpolates between these values.

Essentially, the distribution of $Y_i \mid X_i$ under P_2 has a bump of size ϵ at the boundary of the support of π_b that decays to zero as fast as the Lipschitz assumption will allow.

We now proceed in two steps: first, we verify the separation condition of LeCam's lemma; second, we upper bound the distance between P_1^n and P_2^n.

For the separation condition, we begin by noting that

$$\psi_2^{-,\infty}(P_1) = \mathbb{E}\left[\pi_e(X_i)\left(\sup_{x:\pi_b(x)>0} \mu_1(x) - Ld(X_i, x)\right)\mathbf{1}\{\pi_b(X_i) = 0\}\right], \tag{85}$$

$$= \mathbb{E}\left[\pi_e(X_i)\left(\sup_{x:\pi_b(x)>0} -Ld(X_i, x)\right)\mathbf{1}\{\pi_b(X_i) = 0\}\right]. \tag{86}$$

Suppose that for each X_i, $\sup_{x:\pi_b(x)>0} -Ld(X_i, x)$ is attained at a point $f(X_i)$ and note that $f(X_i)$ must lie on the boundary of the cube $[-1/2, 1/2]^d$ (since it is the projection of X_i onto the support of π_b). Then,

$$\psi_2^{-,\infty}(P_2) = \mathbb{E}\left[\pi_e(X_i)\left(\sup_{x:\pi_b(x)>0} \mu_2(x) - Ld(X_i, x)\right)\mathbf{1}\{\pi_b(X_i) = 0\}\right], \tag{87}$$

$$\geq \mathbb{E}\left[\pi_e(X_i)\left(\mu_2(f(X_i)) - Ld(X_i, f(X_i))\right)\mathbf{1}\{\pi_b(X_i) = 0\}\right], \tag{88}$$

$$= \epsilon\mathbb{E}[\pi_e(X_i)\mathbf{1}\{\pi_b(X_i) = 0\}] + \mathbb{E}\left[\pi_e(x)(-Ld(X_i, f(X_i)))\mathbf{1}\{\pi_b(X_i) = 0\}\right], \tag{89}$$

$$= \epsilon\mathbb{E}[\pi_e(X_i)\mathbf{1}\{\pi_b(X_i) = 0\}] + \psi_2^{-,\infty}(P_1), \tag{90}$$

where the first inequality bounds the supremum by a particular point, the second equality uses the fact that μ_2 is equal to ϵ on the boundary of $[-1/2, 1/2]^d$, and the third uses the fact that f attains the supremum in (86). Then, it follows from (90) that the separation condition is satisfied with $\delta = \epsilon\mathbb{E}[\pi_e(X_i)\mathbf{1}\{\pi_b(X_i) = 0\}]/2$.

It remains to bound $\|P_1^n - P_2^n\|_{\text{TV}}$. By Pinsker's inequality, we have

$$\|P_1^n - P_2^n\|_{\text{TV}} \leq \sqrt{\frac{1}{2}D_{\text{KL}}(P_1^n \| P_2^n)}, \tag{91}$$

$$\leq \sqrt{\frac{n}{2}D_{\text{KL}}(P_1 \| P_2)}, \tag{92}$$

$$\tag{93}$$

Now, letting p_1 and p_2 be the densities of P_1 and P_2 respectively, we have

$$D_{\text{KL}}(P_1 \| P_2) = \int_{[-1,1]^d} \int_{-\infty}^{\infty} \sum_{a\in\{0,1\}} p_1(x, a, y) \log \frac{p_1(x, a, y)}{p_2(x, a, y)}\, dy\, dx. \tag{94}$$

Note that

$$\sum_{a\in\{0,1\}} p_1(x, a, y) \log \frac{p_1(x, a, y)}{p_2(x, a, y)} = p_1(x)(1 - \pi_b(x))p_1(y \mid x) \log \frac{p_1(x)(1 - \pi_b(x))p_1(y \mid x)}{p_2(x)(1 - \pi_b(x))p_2(y \mid x)} \tag{95}$$

$$+ p_1(x)\pi_b(x)p_1(y \mid x) \log \frac{p_1(x)\pi_b(x)p_1(y \mid x)}{p_2(x)\pi_b(x)p_2(y \mid x)}, \tag{96}$$

$$= \frac{1}{2^d} p_1(y \mid x) \log \frac{p_1(y \mid x)}{p_2(y \mid x)}, \tag{97}$$

since $p_1(x) = p_2(x) = 1/2^d$ for all x. Thus, (94) is

$$= \int_{[-1,1]^d} \int_{-\infty}^{\infty} 2^{-d} p_1(y \mid x) \log \frac{p_1(y \mid x)}{p_2(y \mid x)} \, dy \, dx \tag{98}$$

$$= \int_{[-1,1]^d} D_{\mathrm{KL}}(N(0,1) \| N(\mu_2(x), 1)) \, dx, \tag{99}$$

$$= \int_{[-1,1]^d} \frac{1}{2} \mu_2(x)^2 \, dx, \tag{100}$$

$$= \int_{[-1,1]^d} \tilde{\mu}_2(x)^2 \mathbf{1}\{x \in [-1/2 - L\epsilon, 1/2 + L\epsilon]^d \setminus [-1/2 + L\epsilon, 1/2 - L\epsilon]^d\} \, dx, \tag{101}$$

$$\leq \epsilon^2((1 + 2L\epsilon)^d - (1 - 2L\epsilon)^d), \tag{102}$$

$$\leq \epsilon^2 (4L\epsilon)^d, \tag{103}$$

where the last line uses Lemma A.5, proven in Section A.4.

The calculations above show that $\|P_1^n - P_2^n\|_{\mathrm{TV}}^2 \leq \frac{n}{2}(4L)^d \epsilon^{d+2}$. If we set $\epsilon = (2n)^{-1/(d+2)}(4L)^{-d/(d+2)}$, this gives the bound $\|P_1^n - P_2^n\| \leq 1/2$. This choice of ϵ then gives the lower bound

$$\inf_{\hat{\psi}_2^-} \sup_{P \in \mathcal{P}_L^{\mathrm{Lip}}} \mathbb{E}_P \left[(\hat{\psi}_2^- - \psi_2^{-,\infty}(P))^2 \right] \geq \frac{\delta^2}{4}, \tag{104}$$

$$= \frac{1}{16} \mathbb{E}[\pi_e(X_i) \mathbf{1}\{\pi_b(X_i) = 0\}]^2 \epsilon^2, \tag{105}$$

$$= \frac{1}{16} \mathbb{E}[\pi_e(X_i) \mathbf{1}\{\pi_b(X_i) = 0\}]^2 (2n)^{-2/(d+2)} (4L)^{-2/(d+2)}, \tag{106}$$

as long s the condition that $\epsilon < 1/(2L)$ is satisfied. We can check that this condition holds whenever $n \geq 2^{-d+1} L^2$, completing the proof. □

A.4. Technical lemmas

In this section, we collect several lemmas used in the main proofs.

Lemma A.5. *For any $0 < x < 1$, we have that $(1 + x)^d - (1 - x)^d \leq (2x)^d$.*

Proof. We have

$$(1 + x)^d - (1 - x)^d \leq (1 + x)^d = \sum_{k=0}^{d} \binom{d}{k} x^k \leq x^d \sum_{k=0}^{d} \binom{d}{k} = (2x)^d. \tag{107}$$

□

Lemma A.6. *Let x^* and X_{j^*} be as in Lemma A.3 and let b be a lower bound on the density of X_i. Then $\mathbb{E}[d(x^*, X_{j^*})^2] \leq 2(c_d b n)^{-2/d}$, where c_d is the volume of the unit ball in n-dimensions, n is the sample size, and d is the dimension of the covariate space.*

Proof. Let c_d be the volume of the unit ball in d dimensions. We compute

$$\mathbb{E}[d(x^*, X_{j^*})^2] = \int_0^\infty \mathbb{P}(d(x^*, X_{j^*}) \geq \sqrt{t})\, dt, \tag{108}$$

$$= \int_0^\infty \mathbb{P}(d(x^*, X_j) \geq \sqrt{t})^n\, dt, \tag{109}$$

$$\leq \int_0^\infty (1 - c_d t^{d/2} b)_+^n, \tag{110}$$

$$= (c_d b)^{-2/d} \int_0^\infty (1 - u^{d/2})_+^n\, du, \qquad (u = t(c_d b)^{2/d})$$

$$= (c_d b)^{-2/d} \int_0^1 (1 - u^{d/2})^n\, du. \tag{111}$$

If $d = 1$, we can manually compute that (111) is

$$\frac{2(c_d b)^{-2/d}}{n^2 + 3n + 2} \leq \frac{2(c_d b)^{-2/d}}{n^2} = 2(c_d bn)^{-2/d}, \tag{112}$$

and the result of the lemma is satsified. We now assume that $d > 1$, and continue bounding by

$$\leq (c_d b)^{-2/d} \int_0^1 \exp(-u^{d/2} n)\, du, \tag{113}$$

$$\leq (c_d bn)^{-2/d} \int_0^{n^{2/d}} \exp(-v^{d/2})\, dv, , \qquad (v = un^{2/d})$$

$$= (c_d bn)^{-2/d} \int_0^{n^{2/d}} \exp(-v^{d/2}), \tag{114}$$

$$= (c_d bn)^{-2/d} \left(\int_0^1 \exp(-v^{d/2})\, dv + \int_1^{n^{2/d}} \exp(-v^{d/2})\, dv \right), \tag{115}$$

$$\leq (c_d bn)^{-2/d} \left(1 + \int_1^{n^{2/d}} \exp(-v)\, dv \right), \tag{116}$$

$$\leq (c_d bn)^{-2/d}(1 + e^{-1}), \tag{117}$$

$$\leq 2(c_d bn)^{-2/d}. \tag{118}$$

So, in all cases, $\mathbb{E}[d(x^*, X_{j^*})^2] \leq 2(c_d bn)^{-d/2}$. □

B. Computational improvements on $\hat{\psi}_2^-$

In this section, we discuss approximations of

$$\hat{\psi}_2^- = \frac{1}{n} \sum_{i=1}^n \pi_e(X_i) \left(\max_{j : \pi_b(X_j) > 0} \hat{\mu}(X_j) - Ld(X_i, X_j) \right) \mathbf{1}\{\pi_b(X_i) = 0\} \tag{119}$$

with favorable computational properties. The idea is to bound the maximum in (119) by the value of $\hat{\mu}(X_j) - Ld(X_i, X_j)$ for some particular j. A natural choice of j is the index of the nearest neighbor of X_i in the overlap region,

$$\mathrm{nn}(i) = \operatorname*{argmin}_{j : \pi_b(X_j) > 0} d(X_i, X_j). \tag{120}$$

Then, we define

$$\hat{\psi}_2^{-,\mathrm{cons}} = \frac{1}{n} \sum_{i=1}^n \pi_e(X_i) \left(\hat{\mu}(X_{\mathrm{nn}(i)}) - d(X_i, X_{\mathrm{nn}(i)}) \right) \mathbf{1}\{\pi_b(X_i) = 0\}, \tag{121}$$

as a conservative approximation of $\hat{\psi}_2^-$. By conservative, we mean that $\hat{\psi}_2^- \geq \hat{\psi}_2^{-,\text{cons}}$, so that the bounds obtained by using $\hat{\psi}_2^{-,\text{cons}}$ are always wider than those obtained by using $\hat{\psi}_2^-$. This ensures that the interval $[\hat{\psi}^-, \hat{\psi}^+]$ constructed using $\hat{\psi}_2^{-,\text{cons}}$ will still be consistent for $\psi(P_0)$ in the sense of Lemma 4.1.

However, computing $\hat{\psi}_2^{-,\text{cons}}$ will typically be much faster than computing $\hat{\psi}_2^-$. This is because computing $\hat{\psi}_2^{-,\text{cons}}$ only requires finding the nearest overlap neighbor of each point in the no-overlap region, and then evaluating $\hat{\mu}(X_{\text{nn}(i)}) - d(X_i, X_{\text{nn}(i)})$ for that nearest neighbor, rather than evaluating over all points in the overlap region and taking the maximum. This means that, once we have found the nearest overlap neighbor of each point in the no-overlap region, we need only $O(n)$ further computations to compute $\hat{\psi}_2^{-,\text{cons}}$.

We now consider the complexity of nearest-neighbor search problem. To find an exact nearest-neighbor for each point in the no-overlap region generally requires computing all of the pairwise distances $d(X_i, X_j)$, and thus will still take time $O(n^2)$. However, we find in practice that since only $O(n)$ further computation is required to compute $\hat{\psi}_2^{-,\text{cons}}$, this approach is still faster than exactly computing $\hat{\psi}_2^-$. Furthermore, there exist data structures for exact nearest-neighbor search which use various heuristics to improve performance, and these can be leveraged for further computational gains (Bentley, 1975; Omohundro, 1989).

For settings that require a method that is faster than $O(n^2)$, there are two possibilities depending on the dimensionality of the covariates. If the covariates X_i are one-dimensional, methods based on Voronoi diagrams can be used to compute exact nearest neighbors for all points in the overlap region in $O(n)$ time (Har-Peled, 2011; Preparata and Shamos, 2012). In higher dimensions, we can instead use approximate nearest-neighbor search algorithms. For example, if the covariates X_i lie in $\mathbb{R}^p$, and the metric d is Euclidean distance, there exist algorithms that return a point j^* such that $d(X_i, X_{j^*}) \leq cd(X_i, X_{\text{nn}(i)})$ in time $O(n^{1/(2c^2-1-o(1)} + pn^{o(1)}))$ (Andoni et al., 2018). For a moderate value of c, such as $c = 2$, this gives a runtime of $O(n^{1/(7-o(1)} + pn^{o(1)})$ for each point in the no-overlap region. Since there are $O(n)$ points in the no-overlap region, this leads to a total runtime of $O(n^{8/7+o(1)} + pn^{1+o(1)})$, improving significantly on the $O(n^2)$ time required when using exact nearest neighbors. Appealingly, even if we use an approximate nearest neighbor rather than an exact one, we will still obtain a conservative estimate of $\hat{\psi}_2^-$, and thus retain statistical validity.

Finally, another computational advantage of $\hat{\psi}_2^{-,\text{cons}}$ is that once we have identified the nearest-neighbor of each point in the overlap region, we can compute $\hat{\psi}_2^{-,\text{cons}}$ for any value of L with only $O(n)$ operations. This makes search over a large range of values of L, as is done for sensitivity analyses like those shown in Section 6, quite efficient as well.

C. Additional experimental results

C.1. Additional results for Yeast dataset

In Figure 4, we plot the results from Table 1 for $L = 3, 4, 5, \infty$, since these were the values for which we determined (based on the feasibility of the optimization problem) that the smoothness assumption was plausible. We see that, for these values of L, the coverage as defined in Theorem 4.1 with $\epsilon = 0.01$ approaches 100%.

In Table 3, we show the results of Table 1 when defining coverage as in Theorem 4.1 with $\epsilon = 0.005$ rather than $\epsilon = 0.01$, as in the main text. We see that coverage rates are lower, but still approach 1 as the sample size increases. Crucially, the coverage for large values of L remains comparable to the coverage of the Manski intervals, suggesting that the Lipschitz smoothness assumptions for those values of L are plausible.

C.2. Additional results for Yahoo! Front Page Today dataset

In Figure 5, we repeat the experiment on the Yahoo! Front Page Today dataset described in Section 6 for a wider range of smoothness parameters L and cutoffs T. With this range of values, we make two new observations: first, the convergence of our interval to the Manski interval in the upper endpoint was not apparent for small values of L, but is apparent for the values of L considered here. Second, when $T = 0.5$, there are no longer any overlap violations, and our partial identification intervals have length zero and recover the IPW estimator $\hat{\psi}_1$.

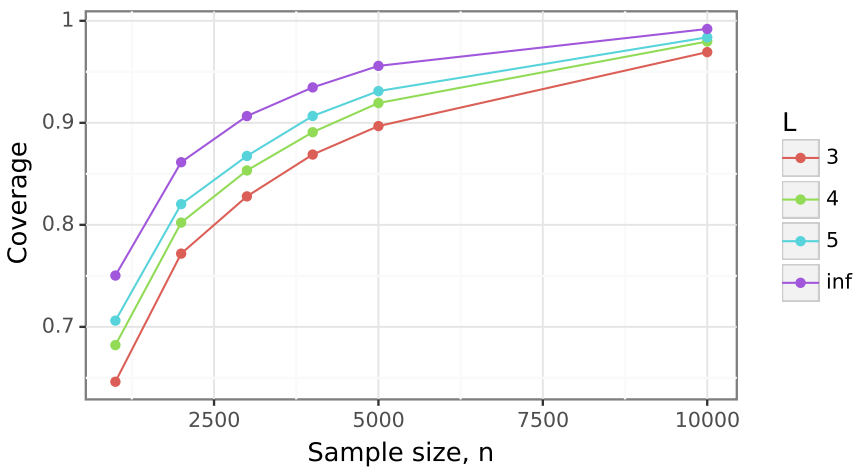

Figure 4. Visualization of results from Table 1 for the values of L for which the optimization problem is consistently feasible and thus the smoothness assumption is plausible. We see that as $n \to \infty$, the coverage (as defined in Theorem 4.1 with $\epsilon = 0.01$) approaches 100%.

Table 3. The same experiment as in Table 1, but with coverage defined using $\epsilon = 0.005$ in Theorem 4.1 rather than $\epsilon = 0.01$. With the smaller value of ϵ, coverage rates are lower, but still approach 1 as the sample size increases. Furthermore, the coverage of the Manski intervals is comparable to the coverage of the Lipschitz intervals for large values of L, indicating that those smoothness assumptions are plausible.

L	$n = 1000$	$n = 2000$	$n = 3000$	$n = 4000$	$n = 5000$	$n = 10000$
1	0.238 (0.785)	0.0002 (0.002)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)
2	0.392 (1.00)	0.493 (1.00)	0.543 (1.00)	0.576 (1.00)	0.603 (0.995)	0.060 (0.106)
3	0.450 (1.00)	0.561 (1.00)	0.613 (1.00)	0.643 (1.00)	0.676 (1.00)	0.766 (1.00)
4	0.498 (1.00)	0.6057 (1.00)	0.657 (1.00)	0.685 (1.00)	0.723 (1.00)	0.810 (1.00)
5	0.528 (1.00)	0.632 (1.00)	0.687 (1.00)	0.716 (1.00)	0.752 (1.00)	0.841 (1.00)
∞	0.591 (1.00)	0.706 (1.00)	0.757 (1.00)	0.789 (1.00)	0.819 (1.00)	0.895 (1.00)

D. Counterexample for no-interaction under smoothness and monotonicity

In this section, we present an example showing that the no-interaction property fails if we make both smoothness and monotonicity assumptions.

To construct our example, we consider a problem with $n = 3$ points with two-dimensional covariates X_i such that $X_1 = (0, 0)$ is the origin, X_2 lies in the first quadrant, and X_3 lies in the fourth quadrant. We assume that X_1 is in the overlap region, and that X_2 and X_3 are not. This configuration is illustrated Figure 6.

We will assume both that the conditional mean μ is L-Lipschitz and that it is monotone with respect to the dictionary order $\prec$. That is, we have $(x_1, y_1) \prec (x_2, y_2)$ if $x_1 \leq x_2$ and $y_1 \leq y_2$.

With this set-up, we have that $t_1 = \hat{\mu}(X_1)$ since X_1 lies in the overlap region. The constraints induced on the t_2 point by t_1 are

$$t_2 \geq \hat{\mu}(X_1) - Ld(X_1, X_2) \quad \text{and} \quad t_2 \geq \hat{\mu}(X_1), \tag{122}$$

since $X_1 \prec X_2$. The latter of these is always tighter, and so to minimize the objective we set $t_2 = \hat{\mu}(X_1)$. With this choice of t_2, the bounds induced on t_3 by t_1 and t_2 are

$$t_3 \geq \hat{\mu}(X_1) - Ld(X_1, X_3) \quad \text{and} \quad t_3 \geq \hat{\mu}(X_1) - Ld(X_2, X_3), \tag{123}$$

respectively. If we select X_2 and X_3 so that $d(X_2, X_3) < d(X_1, X_3)$, then the bound induced on t_3 by t_2 will be tighter than the bound induced on t_3 by t_1, and so we see that the constraints in the no-overlap region do interact with each other in this case.

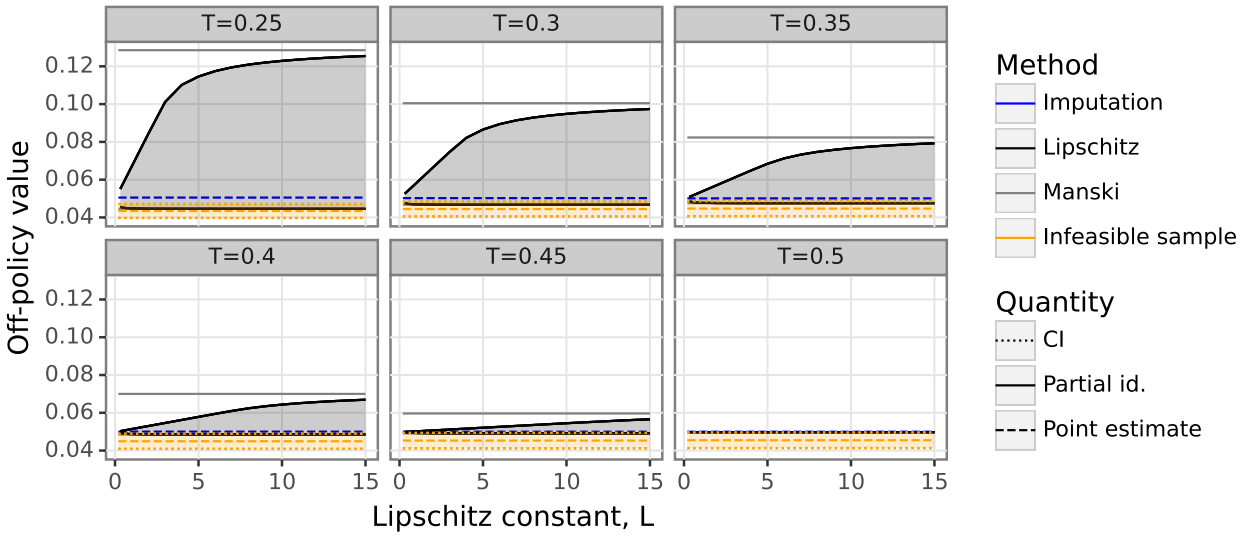

Figure 5. The same experiment as in Figure 3 over a wider range of values of T and L. With this range of values, we see convergence of our bounds to the Manski bounds as L grows large, and also that when $T = 0.5$ and there are no longer any overlap violations, our intervals have width zero as expected.

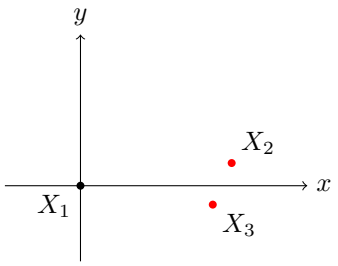

Figure 6. Counterexample showing that the no-interaction condition fails when making both a Lipschitz smoothness assumption and a monotonicity assumption. Here, X_1 lies in the overlap region, but X_2 and X_3 do not, and the only ordering between the points is that $X_1 \prec X_2$. Thus, the tight bound that the $i = 1$ unit implies on the $i = 2$ unit is $t_2 \geq \hat{\mu}(X_1)$, and propagating this to the $i = 3$ unit gives that $t_3 \geq \hat{\mu}(X_1) - Ld(X_2, X_3)$. This will be tighter than the bound directly coming from the $i = 1$ unit, $t_3 \geq \hat{\mu}(X_1) - Ld(X_1, X_3)$, whenever we have $d(X_2, X_3) < d(X_1, X_3)$, as in the figure, and so there is interaction between points in the no-overlap region.

This example thus highlights that our results in Section 4 are non-trivial and highlight special properties of the smoothness assumption that have not been previously observed; further characterization of what kinds of assumptions and combinations satisfy the no-interaction property is an interesting direction for future work.

E. Connections to results of Ben-Michael et al. (2021)

In this section, we describe how our results build on and extend the work of Ben-Michael et al. (2021). In our notation, the procedure of Ben-Michael et al. (2021) for policy learning with no overlap is given by the max-min optimization problem

$$\operatorname*{argmax}_{\pi \in \Pi} \min_{P \in \mathcal{P}_L^{\text{Lip}}} \frac{1}{n} \sum_{i=1}^{n} \mu_P(X_i)\pi(X_i)\mathbf{1}\{\pi_b(X_i) = 0\}, \tag{124}$$

for some policy class Π. To solve this problem, Ben-Michael et al. (2021) observe that

$$\min_{P \in \mathcal{P}_L^{\mathrm{Lip}}} \frac{1}{n} \sum_{i=1}^{n} \mu_P(X_i)\pi(X_i)\mathbf{1}\{\pi_b(X_i) = 0\} \geq \frac{1}{n} \sum_{i=1}^{n} \pi(X_i) \left(\max_{j:\pi_b(X_j)>0} \tilde{\mu}(X_j) - Ld(X_i, X_j) \right) \mathbf{1}\{\pi_b(X_i) = 0\}, \quad (125)$$

where $\tilde{\mu}$ is a simultaneous lower confidence bound on the conditional mean function μ_{P_0}, and then solve the more conservative problem

$$\operatorname*{argmax}_{\pi \in \Pi} \frac{1}{n} \sum_{i=1}^{n} \pi(X_i) \left(\max_{j:\pi_b(X_j)>0} \tilde{\mu}(X_j) - Ld(X_i, X_j) \right) \mathbf{1}\{\pi_b(X_i) = 0\} \quad (126)$$

to learn a policy.

For the purposes of policy learning, such a conservative lower bound is sufficient—indeed, if we were to subtract a large constant, say 100, from the right-hand side of (125), the solution to (126) would remain unchanged. However, for policy evaluation, such conservative bounds are not acceptable: we would like to use the exact value of the left-hand side of (125) as a lower bound, and not incur needlessly wide bounds.

The calculation of this exact value is the contribution of our Theorem 4.5(a), which shows that

$$\min_{P \in \mathcal{P}_L^{\mathrm{Lip}}} \frac{1}{n} \sum_{i=1}^{n} \mu_P(X_i)\pi(X_i)\mathbf{1}\{\pi_b(X_i) = 0\} = \frac{1}{n} \sum_{i=1}^{n} \pi(X_i) \left(\max_{j:\pi_b(X_j)>0} \hat{\mu}(X_j) - Ld(X_i, X_j) \right) \mathbf{1}\{\pi_b(X_i) = 0\}, \quad (127)$$

where $\hat{\mu}$ is an estimate of μ_{P_0}. Thus, the right-hand side of the previous display, which is exactly our $\hat{\psi}_2^-$, is the correct estimator of a bound on the off-policy value. This result, along with its proof based on the no-interaction property, is novel to our work, and guides practitioners on dealing with overlap violations in a policy evaluation problem while ensuring tight bounds.