RIETI Discussion Paper Series 22-E-097

Off-policy Evaluation with General Logging Policies: Implementation at Mercari

NARITA, Yusuke
RIETI

OKUMURA, Kyohei
Northwestern University

SHIMIZU, Akihiro
Mercari

YATA, Kohei
University of Wisconsin-Madison

The Research Institute of Economy, Trade and Industry
https://www.rieti.go.jp/en/

RIETI Discussion Paper Series 22-E-097

October 2022

Off-policy Evaluation with General Logging Policies: Implementation at Mercari[i]

Yusuke NARITA
Yale University and Research Institute of Economy, Trade and Industry
Kyohei OKUMURA
Northwestern University
Akihiro SHIMIZU
Mercari
Kohei YATA
University of Wisconsin-Madison

Abstract

Off-policy evaluation (OPE) attempts to predict the performance of counterfactual policies using log data from a different policy. We extend its applicability by developing an OPE method for a class of both full support and deficient support logging policies in contextual-bandit settings. This class includes deterministic bandit (such as Upper Confidence Bound) as well as deterministic decision-making based on supervised and unsupervised learning. We prove that our method's prediction converges in probability to the true performance of a counterfactual policy as the sample size increases. We validate our method with experiments on partly and entirely deterministic logging policies. Finally, we apply it to evaluate coupon targeting policies by a major online platform and show how to improve the existing policy.

Keywords: Counterfactual Machine Learning, Off-Policy Evaluation, Causal Inference, Regression Discontinuity Design, Coupon Optimization

[i] This study is conducted as a part of research studies at the Research Institute of Economy, Trade and Industry (RIETI).

Off-policy Evaluation with General Logging Policies: Implementation at Mercari

1 Introduction

In bandit and reinforcement learning, off-policy (batch) policy evaluation attempts to estimate the performance of some counterfactual policy given data from a different logging policy. Off-policy evaluation (OPE) is essential when deploying a new policy might be costly or risky, such as in education, medicine, consumer marketing, and robotics. OPE relates to other fields that study counterfactual/causal reasoning, such as statistics and economics.

Most existing OPE studies focus on *full support* logging policies, which take all actions with positive probability in any context, such as stochastic bandit (e.g. ϵ-greedy and Thompson Sampling) and random A/B testing. However, real-world decision-making often uses *deficient support* logging policies, including deterministic bandit (e.g. Upper Confidence Bound) as well as deterministic decision-making based on predictions obtained from supervised and unsupervised learning. An example in the latter group is a policy that greedily chooses the action with the largest predicted reward. OPE is difficult with a deficient support logging policy, since its log data contain no information about the reward from actions never chosen by the logging policy. There appears to be no established OPE estimator for deficient support logging policies (Sachdeva, Su, and Joachims 2020).

We provide a solution to this problem. Our proposed OPE estimator is applicable not only to full support logging policies but also to deficient support ones. We also allow for hybrid stochastic and deterministic logging policies, i.e., logging policies that choose actions stochastically for some individuals and deterministically for other individuals.

Method. Our OPE estimator is based on a modification of the Propensity Score (Rosenbaum and Rubin 1983), which we dub the "Approximate Propensity Score" (APS). APS of action (arm) a at context (covariate) value x is the average probability that the logging policy chooses action a over a shrinking neighborhood around x in the context space. If two actions have nonzero APS at x, the logging policy chooses both actions locally around x. This enables us to estimate the difference in the mean reward between the two actions by exploiting the local subsample around x. When the logging policy is deterministic, the subsample consists of individuals near the decision boundary between the two actions. We then use the estimated reward differences to construct an estimator for the performance of any given counterfactual policy.

As the main theoretical result, we prove that our proposed OPE estimator is consistent. That is, the estimator converges in probability to the true performance of a counterfactual policy as the sample size increases, under the assumption that the mean reward differences are constant over the context space (Theorem 7). This result holds whether the logging policy is of full support or deficient support. The proof exploits results from differential geometry and geometric measure theory, which have not been applied in machine learning research as far as we know.

Simulation Experiments. We validate our method with two simulation experiments. The first considers a mix of full support and deficient support policies as the logging policy. Actions are randomly chosen for a small A/B test segment of the population and are chosen by a deterministic supervised learning algorithm for the rest of the population. For the task of evaluating counterfactual policies, our method produces smaller mean squared errors than a baseline estimator that only uses the A/B test subsample. The second experiment considers a situation in which we have a batch of data generated by a deterministic bandit algorithm. We find that our estimator outperforms a regression-based estimator in terms of mean squared errors.

Real-World Application. We empirically apply our method to evaluate and optimize coupon targeting policies. Our application is based on proprietary data provided by Mercari Inc., a major e-commerce company running online C2C marketplaces in Japan and the US. This company uses a deterministic policy based on uplift modeling to decide whether they offer a promotional coupon to each target customer. We use the data produced by their policy and our method to evaluate a counterfactual policy that offers the coupon to more customers. Our method predicts that the counterfactual policy would increase revenue more than the cost of coupon offers, suggesting that redesigning the current policy is profitable.

Related Work. Widely-used OPE methods include inverse probability weighting (IPW) (Precup 2000; Strehl et al. 2010), self-normalized IPW (Swaminathan and Joachims 2015), Doubly Robust (Dudík et al. 2014), and more advanced variants (Wager and Athey 2018; Farajtabar, Chow, and Ghavamzadeh 2018; Su et al. 2020). These methods are based on importance sampling (IS) and require that the logging policy is of full support, i.e., assigns a positive probability to every action potentially chosen by the counterfactual policy. This restriction makes them hard to use when the logging policy is of deficient support.

There are two existing approaches to deficient support logging policies. [1] The first approach considers a logging policy that varies over time or across individuals (Strehl et al. 2010). Viewing the sequence of varying logging policies as a single full support logging policy, it is possible to apply IS-based OPE methods. Unlike this approach, our approach is usable even when the logging policy is fixed.

The second approach, called the Direct Method or Regression Estimator, predicts the mean reward conditional on the action and context by supervised learning and uses the prediction to estimate the performance of a counterfactual policy (Beygelzimer and Langford 2009; Dudík et al. 2014). Similar regression-based methods are proposed for reinforcement learning settings (Duan, Jia, and Wang 2020). This approach is sensitive to the accuracy of the mean reward prediction. It may have a large bias if the regression model is not correctly specified. This issue is particularly severe when the logging policy is of deficient support, since each action is observed only in a limited area of the context space. Our approach instead predicts the mean reward differences between actions by exploiting local subsamples near the decision boundaries without specifying the regression model. This idea relates to regression discontinuity designs in the social sciences (Lee and Lemieux 2010).

It is worth noting that our approach is applicable to *off-policy selection*, in which the researcher is to design a decision rule to select a policy given a finite set of policies (Kuzborskij et al. 2021). Since our method can estimate the expected reward of the policies, we can first estimate the rewards of the policies, and then choose the one with the highest expected reward.

2 Framework

$\mathcal{A} \coloneqq \{1, ..., m\}$ is a set of *actions* that the decision maker can choose from. Let $\mathbb{R}^p$-valued random variable X denote the *context* that the decision maker observes when picking an action. Let $\mathcal{X}$ denote the support of X. To simplify the exposition, we assume that X is continuously distributed. Let a tuple of m $\mathbb{R}$-valued random variables $(Y(1), \dots, Y(m))$ denote *potential rewards*; $Y(a)$ denotes a *potential reward* that is observed when action a is chosen. $(Y(1), \dots, Y(m), X)$ follows distribution P, which is unknown to the decision maker.

A *policy* chooses an action given a context. Let $ML : \mathbb{R}^p \to \Delta(\mathcal{A})$ represent the *logging policy*, where $ML(a|x)$ is the probability of taking action a for individuals with context x. We assume that the analyst knows the logging policy and is able to simulate it. That is, the analyst is able to compute the probability $ML(a|x)$ for each action $a \in \mathcal{A}$ given any context $x \in \mathbb{R}^p$. Suppose we have log data $\{(Y_i, X_i, A_i)\}_{i=1}^n$ generated as follows. For each individual i, (1) $(Y_i(1), \dots, Y_i(m), X_i)$ is i.i.d. drawn from P; [2] (2) Given X_i, the action A_i is randomly chosen based on the probability $ML(\cdot|X_i)$; (3) We observe the reward $Y_i \coloneqq Y_i(A_i)$. Note that only one of $Y_i(1), \dots, Y_i(m)$ is observed for individual i and recorded as Y_i in the log data. The joint distribution of (Y, X, A) is determined once ML and P are given.

Prediction Target. We are interested in estimating the expected reward from any given *counterfactual policy* $\pi : \mathbb{R}^p \to \Delta(\mathcal{A})$, which chooses a distribution of actions given individual context:

$$V(\pi) \equiv E\left[\sum_{a \in \mathcal{A}} Y(a)\pi(a|X)\right].$$

3 Learning with Infinite Data

We first consider the identification problem, which asks whether it is possible to learn $V(\pi)$ if we had an infinite amount of data. Formally, we say that $V(\pi)$ is *identified* if it is uniquely determined by the joint distribution of (Y, X, A). A key step toward answering the identification question is what we call the *Approximate Propensity Score* (APS). To define it, for $a \in \mathcal{A}$ and $x \in \mathcal{X}$, let:

$$p_\delta^{ML}(a|x) \equiv \frac{\int_{B(x,\delta)} ML(a|x^*)dx^*}{\int_{B(x,\delta)} dx^*},$$

where $B(x, \delta) = \{x^* \in \mathbb{R}^p : \|x - x^*\| < \delta\}$ is the δ-ball around $x \in \mathcal{X}$. Here, $\|\cdot\|$ denotes the Euclidean distance on $\mathbb{R}^p$. To make common δ for all dimensions reasonable, we normalize X_{ij} to have mean zero and variance one for each $j = 1, ..., p$. We assume that ML is a Lebesgue measurable function so that the integrals exist. We then define APS p^{ML} as follows: for $a \in \mathcal{A}$ and $x \in \mathcal{X}$,

$$p^{ML}(a|x) \equiv \lim_{\delta \to 0} p_\delta^{ML}(a|x).$$

Figure 1 illustrates APS. Here $X \in \mathbb{R}^2$, $\mathcal{A} = \{1, 2, 3\}$, and the support of X is divided into four sets depending on the value of ML as in panel (a). Panel (b) shows the corresponding APS. For the interior points of each of the four sets, APS is equal to ML. On the border of any two sets, APS is the average of the ML values in the two sets.

Our identification analysis uses the following assumption.

Assumption 1 (Local Mean Continuity)**.** For any $a \in \mathcal{A}$, the conditional expectation function $E[Y(a)|X]$ is continuous at each $x \in \mathcal{X}$ such that $p^{ML}(a|x) > 0$ and $ML(a|x) = 0$.

$ML(a|x) = 0$ means that action a is never taken for individuals with context x. If APS of a at x is nonzero ($p^{ML}(a|x) > 0$), however, there exists a point close to x that has a positive probability of receiving action a, which enables us to observe the reward from the action near x. For any such point x, Assumption 1 ensures that the points close to x have similar conditional means of the potential reward $Y(a)$. Thus, the conditional mean reward from action a at x

[1] Sachdeva et al. (2020) also proposes another approach in which they restrict the policy space.

[2] This assumption is valid when we have a batch of log data generated by a fixed policy.

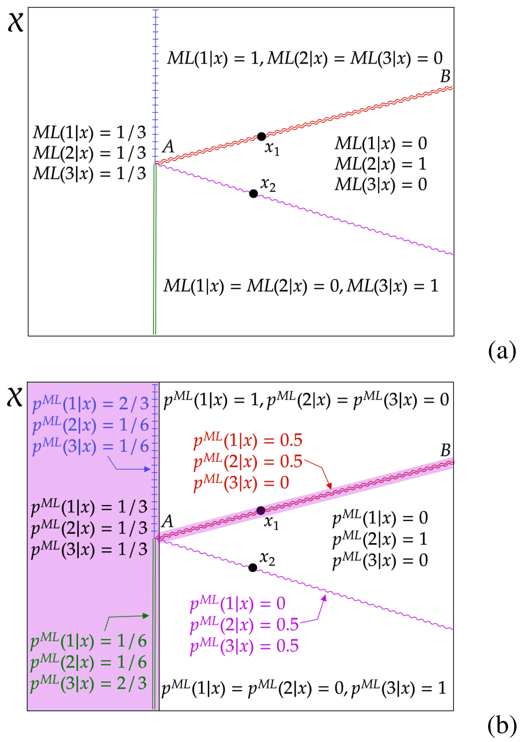

Figure 1: Example of the Approximate Propensity Score

Notes: This figure shows an example of logging policy ML (panel (a)) and corresponding APS p^{ML} (panel (b)). The shaded region in panel (b) indicates the subpopulation for which $p^{ML}(1|x) > 0$ and $p^{ML}(2|x) > 0$. As discussed in Section 4, our method uses the subsample in the shaded region to estimate the conditional mean difference $E[Y(2)|X] - E[Y(1)|X]$.

is identified. On the other hand, when $ML(a|x) > 0$, action-context pair (a, x) is observed, allowing us to identify the mean reward without any assumptions. Assumption 1 therefore does not impose continuity at such points. The lemma below summarizes the above argument. For a set $A \subset \mathbb{R}^p$, let $\mathrm{int}(A)$ denote the interior of A.

Lemma 2 (Identification of Conditional Means)**.** *If Assumption 1 holds, then for each $a \in \mathcal{A}$, $E[Y(a)|X = x]$ is identified for every $x \in \mathrm{int}(\mathcal{X})$ such that $p^{ML}(a|x) > 0$.*

We use Lemma 2 to analyse identification of $V(\pi)$. Suppose first that $\pi(a|x) > 0 \Longrightarrow p^{ML}(a|x) > 0$, that is, the counterfactual policy π only chooses actions with nonzero APS. Lemma 2 implies that the conditional mean reward is identified at every (a, x) pair that could be realized under the policy π. As a result, the expected reward $V(\pi)$ is identified for any such policy. However, if there exists (a, x) such that $\pi(a|x) > 0$ but $p^{ML}(a|x) = 0$, we cannot identify $V(\pi)$ without additional assumptions. To be able to identify $V(\pi)$ for any policy π, we assume that the difference in the conditional mean reward function $E[Y(a)|X]$ between any two actions is constant over $x \in \mathcal{X}$.

Assumption 3 (Constant Conditional Mean Differences)**.** There exists a function $\beta : \mathcal{A} \times \mathcal{A} \to \mathbb{R}$ such that $E[Y(a)|X] - E[Y(a')|X] = \beta(a, a')$.

Appendix B includes discussion about what would happen if we drop Assumption 3 and a potential way of relaxing this. We also impose the following condition on APS.

Assumption 4 (Existence of Nonzero APS)**.** For every $a \in \{2, ..., m\}$, there exists a sequence $\{a_1, ..., a_L\}$ with $a_1 = 1$ and $a_L = a$ for which the following condition holds: for every $l \in \{1, ..., L-1\}$, there exists $x \in \mathrm{int}(\mathcal{X})$ such that $p^{ML}(a_l|x) > 0$ and $p^{ML}(a_{l+1}|x) > 0$.

Assumption 4 states that there exists a path from a baseline action ($a_1 = 1$) to any other action ($a_L = a$) for which APS of any two consecutive actions (a_l and a_{l+1}) is positive at some x. For example, suppose that $m = 3$, $p^{ML}(1|x_1) > 0$, $p^{ML}(2|x_1) > 0$, $p^{ML}(2|x_2) > 0$ and $p^{ML}(3|x_2) > 0$ for some $x_1, x_2 \in \mathcal{X}$ as in Figure 1 (b). In this case, the sequence $\{1, 2\}$ satisfies the condition in Assumption 4 for $a = 2$, and the sequence $\{1, 2, 3\}$ satisfies the condition for $a = 3$. By Lemma 2, the four conditional means $E[Y(1)|X = x_1]$, $E[Y(2)|X = x_1]$, $E[Y(2)|X = x_2]$ and $E[Y(3)|X = x_2]$ are identified. Hence, the two differences $E[Y(1)|X = x_1] - E[Y(2)|X = x_1]$ and $E[Y(2)|X = x_2] - E[Y(3)|X = x_2]$ are identified. Under Assumption 3, the two differences do not depend on x. As a result, $E[Y(1)|X = x] - E[Y(2)|X = x]$ and $E[Y(2)|X = x] - E[Y(3)|X = x]$ are identified for every $x \in \mathcal{X}$. Noting that $E[Y(a)|X = x]$ is identified for at least one $a \in \mathcal{A}$ for every $x \in \mathcal{X}$, we can use the differences to identify $E[Y(a)|X = x]$ for every (a, x) pair, even for those not observed in data. Thus, $V(\pi)$ is identified for any policy π.

Proposition 5 (Identification of $V(\pi)$)**.** *Under Assumptions 1–4, $V(\pi)$ is identified for any policy π.*

4 Learning with Finite Data

OPE Estimator. Suppose that we observe a sample $\{(Y_i, X_i, A_i)\}_{i=1}^n$ of size n. We propose an OPE estimator based on the following expression of our prediction target $V(\pi)$: under Assumption 3,

$$V(\pi) = V(ML) + E\left[\sum_{a=2}^{m} \beta(a, 1)\big(\pi(a|X) - ML(a|X)\big)\right]. \tag{1}$$

Appendix G derives this expression. Since $V(ML)$ is the value from the logging policy ML, $V(ML)$ can be estimated by the sample mean of Y_i. Our identification analysis suggests a way of conducting OPE on any policy π: (1) estimate $\beta(a, a')$ for each (a, a') pair such that $p^{ML}(a|x) > 0$ and $p^{ML}(a'|x) > 0$ for some x; (2) use the estimates to recover $\beta(a, 1)$ for every $a \in \{2, ...m\}$ and plug them into the sample analogue of the above expression. For simplicity, we consider a setup in which $p^{ML}(a|x) > 0$ and $p^{ML}(1|x) > 0$ for some x for every a so that we can directly estimate $\beta(a, 1)$ in step (1) above.

To estimate $\beta(a, 1)$, we use the subsample $\mathcal{I}(a; \delta_n) \equiv \{i : A_i \in \{1, a\}, q^{ML}_{\delta_n}(a|X_i) \in (0, 1)\}$, where $q^{ML}_{\delta_n}(a|X_i) \equiv \frac{p^{ML}_{\delta_n}(a|X_i)}{p^{ML}_{\delta_n}(a|X_i) + p^{ML}_{\delta_n}(1|X_i)}$, and δ_n is a given bandwidth. The bandwidth shrinks towards zero as the sample size n increases.[3] $q^{ML}_{\delta_n}(a|X_i)$ can be viewed as APS of action a within the subsample for which either action 1 or a is assigned. The subsample $\mathcal{I}(a;\delta_n)$ contains all observations i such that both actions 1 and a can be chosen by the logging policy locally around X_i. For example, in Figure 1 (b), the shaded region corresponds to the subsample $\mathcal{I}(2;\delta_n)$. This covers not only the subsample subject to full randomization (for which $ML(1|x) = ML(2|x) = ML(3|x) = 1/3$) but also the local subsample near the deterministic decision boundary AB between actions 1 and 2.

We propose minimizing the sum of squared errors on the subsample $\mathcal{I}(a;\delta_n)$:

$$\begin{aligned}&(\hat{\alpha}_a, \hat{\beta}_a, \hat{\gamma}_a) \\ &= \underset{(\alpha_a,\beta_a,\gamma_a)}{\operatorname{argmin}} \sum_{i\in\mathcal{I}(a;\delta_n)} \left(Y_i - \alpha_a - \beta_a 1\{A_i = a\} - \gamma_a q^{ML}_{\delta_n}(a|X_i)\right)^2,\end{aligned} \tag{2}$$

where $1\{\cdot\}$ is the indicator function. $\hat{\beta}_a$ is our estimator of $\beta(a,1)$. We include $q^{ML}_{\delta_n}(a|X_i)$ as an explanatory variable to adjust for imbalance in the context distribution between actions 1 and a, as is done with the standard propensity score (Angrist and Pischke 2008; Hull 2018). We then define our OPE estimator as:

$$\hat{V}(\pi) = \frac{1}{n}\sum_{i=1}^{n}\left(Y_i + \sum_{a=2}^{m}\hat{\beta}_a\big(\pi(a|X_i) - ML(a|X_i)\big)\right). \tag{3}$$

It is worth noting that our method does not require the model selection.

For estimating $\beta(a,1)$, the above method uses APS $p^{ML}_{\delta_n}(a|X_i)$, which may be difficult to compute analytically if ML is complex. In such a case, we propose approximating it by brute force simulation. We draw a value of x from the uniform distribution on $B(X_i,\delta_n)$ a number of times, compute $ML(a|x)$ for each draw, and take the average of $ML(a|x)$ over the draws.[4] We then use it instead of $p^{ML}_{\delta_n}(a|X_i)$ to compute $q^{ML}_{\delta_n}(a|X_i)$, and then compute $\hat{\beta}(a,1)$ and $\hat{V}(\pi)$ as in (2) and (3).

Consistency. We show that $\hat{V}(\pi)$ is a consistent estimator of $V(\pi)$, that is, $\hat{V}(\pi)$ converges in probability to $V(\pi)$ as $n\to\infty$. Our consistency result uses the following assumptions for the subsample assigned to one of the actions a and 1, for every $a \in \{2,...,m\}$. Let $\mathcal{X}_{a,1} \equiv \{x \in \mathcal{X} : ML(a|x) > 0 \text{ or } ML(1|x) > 0\}$, $\widetilde{ML}(a|x) \equiv \Pr(A_i = a|A_i \in \{1,a\}, X_i = x) = \frac{ML(a|x)}{ML(a|x)+ML(1|x)}$, $\mathcal{X}^a_{a,1} \equiv \{x \in \mathcal{X} : \widetilde{ML}(a|x) = 1\}$, and $\mathcal{X}^1_{a,1} \equiv \{x \in \mathcal{X} : \widetilde{ML}(a|x) = 0\}$. In other words, $\mathcal{X}_{a,1}$ is the set of context values for which action 1 or a can be taken, $\widetilde{ML}(a|x)$ is the probability of choosing action a conditional on $A_i \in \{1,a\}$ and $X_i = x$, and $\mathcal{X}^a_{a,1}$ and $\mathcal{X}^1_{a,1}$ are the set of context values for which the conditional probability is 1 and 0, respectively.

Assumption 6. The following holds for all $a \in \{2,...,m\}$.

(a) *(Existence of Subsample)* $\Pr(A_i \in \{1,a\}) > 0$.

(b) (Almost Everywhere Continuity of ML) $ML(a|\cdot)$ and $ML(1|\cdot)$ are continuous almost everywhere on $\mathcal{X}_{a,1}$ with respect to the Lebesgue measure.

(c) (Measure Zero Boundaries of $\mathcal{X}^a_{a,1}$ and $\mathcal{X}^1_{a,1}$). For $a' \in \{1,a\}$, $\mathcal{L}^p(\mathcal{X}^{a'}_{a,1}) = \mathcal{L}^p(\mathrm{int}(\mathcal{X}^{a'}_{a,1}))$, where $\mathcal{L}^p$ is the Lebesgue measure on $\mathbb{R}^p$.

(d) *(Finite Moments)* $E[Y_i^2] < \infty$.

(e) *(Nonzero Conditional Variance)* If $\Pr(\widetilde{ML}(a|X_i) \in (0,1)|A_i \in \{1,a\}) > 0$, then $\mathrm{Var}(\widetilde{ML}(a|X_i)|\widetilde{ML}(a|X_i) \in (0,1), A_i \in \{1,a\}) > 0$.

If $\Pr(\widetilde{ML}(a|X_i) \in (0,1)|A_i \in \{1,a\}) = 0$, then the following conditions (f)–(i) additionally hold.

(f) (Deterministic ML) For all $x \in \mathbb{R}^p$, either $ML(a|x) = 1$ or $ML(a|x) = 0$.

(g) (C^2 Boundary of Ω^*_a) There exists a partition $\{\Omega^*_{a,1},...,\Omega^*_{a,K}\}$ of $\Omega^*_a = \{x \in \mathbb{R}^p : ML(a|x) = 1\}$ (the set of the context values for which the probability of choosing action a is one) such that
   (1) $\mathrm{dist}(\Omega^*_{a,k},\Omega^*_{a,l}) > 0$ for any $k,l \in \{1,...,K\}$ such that $k \neq l$. Here $\mathrm{dist}(S,T) = \inf_{x\in S,y\in T}\|x-y\|$ is the distance between two sets S and $T \subset \mathbb{R}^p$;
   (2) $\Omega^*_{a,k}$ is nonempty, bounded, open, connected and twice continuously differentiable for each $k \in \{1,...,K\}$. [5]

(h) (Regularity of Deterministic ML)
   (1) $\mathcal{H}^{p-1}(\partial\Omega^*_a) < \infty$, $\int_{\partial\Omega^*_a\cap\partial\mathcal{X}_{a,1}} d\mathcal{H}^{p-1}(x) = 0$, and $\int_{\partial\Omega^*_a\cap\mathcal{X}_{a,1}} f_X(x)d\mathcal{H}^{p-1}(x) > 0$, where ∂S denotes the boundary of a set $S \subset \mathbb{R}^p$, f_X is the probability density function of X_i, and $\mathcal{H}^k$ is the k-dimensional Hausdorff measure on $\mathbb{R}^p$. [6]
   (2) There exists $\delta > 0$ such that $ML(a|x) = 1$ or $ML(1|x) = 1$ for almost every $x \in N(\mathcal{X}_{a,1},\delta) \cap N(\partial\Omega^*_a,\delta)$, where $N(S,\delta) = \{x \in \mathbb{R}^p : \|x-y\| < \delta \text{ for some } y \in S\}$ for a set $S \subset \mathbb{R}^p$ and $\delta > 0$.

(i) (Conditional Moments and Density near $\partial\Omega^*_a$) There exists $\delta > 0$ such that
   (1) $E[Y_i(a)|X_i]$, $E[Y_i(1)|X_i]$, and f_X are continuous and bounded on $N(\partial\Omega^*_a,\delta)$;
   (2) $E[Y_i(a)^2|X_i]$ and $E[Y_i(1)^2|X_i]$ are bounded on $N(\partial\Omega^*_a,\delta)$.

Here we only discuss a few key assumptions. Appendix C provides discussion about other assumptions. Note first that Assumption 6 (b) allows the function ML to be discontinuous on a set of points with the Lebesgue measure zero. For example, ML is allowed to be a step function.

[3] For the bandwidth δ_n, we suggest considering several different values and check if the estimates are robust to bandwidth changes. It is hard to pick δ_n in a data-driven way to minimize the mean squared error, since it would require nonparametric estimation of functions on the high-dimensional context space.

[4] The approximation error of the simulated APS relative to true $p^{ML}_{\delta_n}(a|X_i)$ has a $1/\sqrt{S}$ rate of convergence, where S is the number of simulation draws. This rate does not depend on the dimension of X_i, so the simulation error can be made negligible by using a large number of simulation draws even when X_i is high dimensional.

[5] See the Appendix A for definition.

[6] See the Appendix A for definition.

When ML is deterministic, $\partial\Omega_a^*$ corresponds to the decision boundary for action a in the context space. Assumption 6 (g) imposes differentiability of the boundary. The condition is satisfied if, for example, $\Omega_a^* = \{x \in \mathbb{R}^p : f(x) \geq 0\}$ for some twice continuously differentiable function $f : \mathbb{R}^p \to \mathbb{R}$ such that the gradient $\nabla f(x)$ is nonzero for all $x \in \mathbb{R}^p$ with $f(x) = 0$. Furthermore, Assumption 6 (h) (1) assumes that $\partial\Omega_a^*$ is $(p-1)$ dimensional and has nonzero density.

Theorem 7 (Consistency of $\hat{V}(\pi)$)**.** *Suppose that Assumptions 3 and 6 hold, $\delta_n \to 0$, and $n\delta_n \to \infty$ as $n \to \infty$. Then $\hat{V}(\pi)$ converges in probability to $V(\pi)$ for every policy π.*

Our consistency result requires that δ_n goes to zero slower than n^{-1}. This ensures that, when ML is deterministic, we have sufficiently many observations in the δ_n-neighborhood of the boundary of Ω_a^*. Importantly, the rate condition does not depend on the dimension of X_i. This is because we use all the observations in the δ_n-neighborhood of the boundary, and the number of those observations is of order $n\delta_n$ regardless of the dimension of X_i if the boundary is $(p-1)$ dimensional. Our estimator is therefore expected to perform well even if X_i is high dimensional.

Our result holds under the assumption of constant conditional mean reward differences. If this assumption does not hold for a deterministic logging policy, $\hat{\beta}_a$ is a consistent estimator of the mean reward difference for the subpopulation on the decision boundary between actions a and 1 (see Appendix G). Therefore, our estimator may still perform well when we are interested in a counterfactual policy that marginally changes the logging policy's decision boundary.

5 Simulations

Experiment 1: Mix of A/B Test and Deterministic Logging Policy

Consider a tech company that conducts an A/B test using a small segment of the population. The company applies a deterministic logging policy to the rest of the population. We generate a random sample $\{(Y_i, X_i, A_i)\}_{i=1}^n$ of size n = 50,000 as follows. There are 5 actions ($m = 5$) and 100 context variables ($p = 100$), with $X_i \sim N(0, \mathbf{\Sigma})$. $Y_i(a)$ is generated as $Y_i(a) = 0.75\sum_{k=1}^{100} X_{ki}^2\alpha_{0,k} + 0.25u_i + \epsilon_i(a)$, where $\alpha_0 = (\alpha_{0,1}, ..., \alpha_{0,100}) \in \mathbb{R}^{100}$, $u_i \sim N(0, 1)$, and $\epsilon_i(a) \sim N(a, 1)$. The conditional mean difference $E[Y_i(a)|X_i] - E[Y_i(1)|X_i]$ is constant over x. The choice of parameters $\mathbf{\Sigma}$ and α_0 is explained in Appendix D. To generate A_i, let $q_{0.99}^k$ be the 99th percentile of the kth context variable X_{ki}. Let $\tau_{pred}^{ML}(x, a)$ be a prediction of the reward from action a given context value x obtained by supervised learning from a past, independent training sample $\tilde{\mathcal{D}} = \{(\tilde{Y}_i, \tilde{X}_i, \tilde{A}_i)\}_{i=1}^{\tilde{n}}$ of size $\tilde{n}$ = 10,000 (see Appendix D for how we constructed $\tilde{\mathcal{D}}$ and τ_{pred}^{ML}). A_i is then generated based on the logging policy:

$$ML(a|x) = \begin{cases} 1/5 & \text{if } x_1 \geq q_{0.99}^1 \\ 1\{a = \underset{a' \in \{1,\ldots,5\}}{\operatorname{argmax}}\ \tau_{pred}^{ML}(x, a')\} & \text{if } x_1 < q_{0.99}^1. \end{cases}$$

The first case corresponds to the A/B test segment while the second case to the deterministic policy segment. Finally, Y_i is generated as $Y_i = Y_i(A_i)$.

We simulate 1,000 hypothetical samples from the above data-generating process. For each simulation, we use the simulated sample to estimate the value of a counterfactual policy π, another mix of an A/B test and a deterministic policy. With another reward prediction function τ_{pred}^{π},

$$\pi(a|x) = \begin{cases} 1/5 & \text{if } x_2 \geq q_{0.99}^2 \\ 1\{a = \underset{a' \in \{1,\ldots,5\}}{\operatorname{argmax}}\ \tau_{pred}^{\pi}(x, a')\} & \text{if } x_2 < q_{0.99}^2. \end{cases}$$

Alternative Methods. We compare our method with two alternative estimators. The first uses the A/B test segment (for which $ML(a|X_i) = 1/5$) while the second uses the full sample. The methods first compute the simple mean differences in reward Y_i between actions $a \in \{2, ..., 5\}$ and 1, and then plugs them into $\hat{\beta}_a$ of Eq. (3). Both our method and the alternative estimator with the A/B test segment produce consistent estimators of the prediction target $V(\pi)$. However, the alternative uses only the A/B test segment while our method additionally uses the local subsample near the decision boundary of the deterministic policy as we discussed in Section 4.

Result. The first panel of Table 1 presents the bias, standard deviation (S.D.) and root mean squared error (RMSE) of our proposed estimators with several choices of δ and two alternative estimators. The alternative estimator using the full sample has a larger bias than the other two, since it does not control for the difference in the context distribution between actions. Our proposed estimator outperforms the alternative estimator using the A/B test sample in terms of RMSE. This suggests that exploiting both of the A/B test segment and the local subsample near the deterministic decision boundary can lead to better performance than using only the A/B test segment.

Experiment 2: Upper Confidence Bound Logging Policy

In the second experiment, both the logging policy and the counterfactual policy are deterministic. The rest of the setup is the same as that in the first experiment. We first use the independent training sample $\tilde{\mathcal{D}}$ to train an Upper Confidence Bound bandit algorithm. The logging policy ML is given by $ML(a|x) = 1\left\{a = \arg\max_{a' \in \{1,\ldots,5\}} UCB(x, a')\right\}$, where $UCB(x, a)$ is an upper confidence bound of $E[Y_i(a)|X_i = x]$. See Appendix D for training details. We do not update the policy while generating $\{(Y_i, X_i, A_i)\}_{i=1}^n$ in the simulation. The sample is a batch of log data.

For the counterfactual policy π, we use $\tilde{\mathcal{D}}$ to train a model $f(x, a)$ that predicts the reward given the context and action, using sklearn's RandomForestRegressor with 500 trees and otherwise default parameters. The counterfactual policy tries to maximize the expected reward $V(\pi)$ by choosing the action with the largest predicted reward: $\pi(a|x) = 1\left\{a = \arg\max_{a' \in \{1,\ldots,5\}} f(x, a')\right\}$.

Alternative Method. We compare our method with an alternative estimator using the Direct Method. This first fits a linear model $Y_i = \alpha + \sum_{a=2}^{5} \beta_a 1\{A_i = a\} +$

Table 1: Simulation results: bias, S.D., and RMSE of estimators of $V(\pi)$

	Our Proposed Method with APS Controls				Method with Mean Differences		Direct
	$\delta = 0.1$	$\delta = 0.5$	$\delta = 1$	$\delta = 2.5$	A/B Test Sample	Full Sample	Method
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
			Experiment 1: Mix of A/B Test and Deterministic Logging Policy				
Bias	$-.060$	$-.057$	$-.057$	$-.060$	$-.061$	$-.075$	—
S.D.	.099	.098	.096	.096	.101	.103	—
RMSE	.115	.113	.112	.113	.118	.128	—
Avg. N	1862	6362	12502	33122	500	50000	—
			Experiment 2: Upper Confidence Bound Logging Policy				
Bias	.048	.047	.046	.047	—	—	.342
S.D.	.033	.030	.029	.029	—	—	.012
RMSE	.058	.056	.055	.055	—	—	.342
Avg. N	3397	17344	31107	47601	—	—	50000

Notes: This table shows the bias, the standard deviation (S.D.), and the root mean squared error (RMSE) of the estimators of the reward from the counterfactual policy $V(\pi)$ in the two simulation experiments. We use $1,000$ simulations of a size $50,000$ sample to compute these statistics. Columns (1)–(4) report estimates from our method with several choices of δ. Each APS is computed by averaging 100 simulation draws of the ML value. In columns (5)–(6), we estimate the mean reward differences $\beta(a,1)$ by the sample mean differences in the A/B test segment and the full sample, respectively. In column (7), we estimate $\beta(a,1)$ by fitting a linear model that predicts the reward from the context and action. The bottom row of each panel shows the average number of observations with nonzero APS for every action (Columns (1)–(4)), that with nonzero ML for every action (Column (5)), or the total sample size (Columns (6)–(7)).

$\sum_{k=1}^{100} X_{ki}\gamma_k + e_i$, then makes the reward prediction from action a for individual i by $\hat{\mu}_i(a) = Y_i + (\hat{\beta}_a - \hat{\beta}_{A_i})$, and finally computes $\hat{V}(\pi) = \frac{1}{n}\sum_{i=1}^{n}\sum_{a=1}^{5}\hat{\mu}_i(a)\pi(a|X_i)$. The linear model used by this method correctly imposes the constant conditional mean differences but misspecifies the functional form with respect to X_i.

Result. The second panel of Table 1 shows the result. The alternative using the Direct Method is significantly biased due to model misspecification. Our proposed estimator seems to effectively use the local subsample near the decision boundary and has smaller bias and RMSE than the alternative.

6 Real-World Application

Setup. We apply our method to empirically evaluate a coupon targeting policy of an online platform. This application uses proprietary data provided by Mercari, Inc.. This company conducts the following promotional campaign. They target customers who signed up for Mercari 4 days ago but have not made a purchase yet. The company uses a logging policy based on an uplift model to determine whether they offer a promotional coupon to each target customer. If customers receive the coupon and make a purchase, they get 900 points (equivalent to 8.34 USD) that they can use for future purchases. We observe data (Y_i, X_i, A_i) for each target user i from this campaign, where action $A_i \in \{0,1\}$ is whether the logging policy recommended offering the coupon to the customer ($A_i = 1$) or not ($A_i = 0$), X_i is the vector of more than 200 input features for the uplift model, and Y_i is an outcome such as the customer's spending after this coupon offer.

The company's logging policy works as follows. They first use data from a past A/B test and XGBoost to train a model of the conditional average effect of the coupon on purchases (they use library pylift for implementation). Let $\tau(x)$ be the predicted coupon effect for those whose feature value is $X_i = x$. The logging policy then recommends offering a coupon to customer i if the predicted effect is in the top 80% of the distribution of predicted effects. That is, the logging policy ML is given by $ML(1|x) = 1\{\tau(x) \geq c\}$, where c is the 20th quantile of the distribution of $\tau(X_i)$.

Effects of Policy Recommendation. We first apply our method to the logged data generated by the above policy to estimate the effect of the policy recommendation A_i ($\beta(1,0) = E[Y_i(1) - Y_i(0)]$) on the following three outcomes: (1) the purchase value (how much the customer spent), (2) the number of transactions, and (3) point usage (how many points the customer used). All outcomes are sums over 18 days after the coupon offer decision. We compute APS with $\delta \in \{0.4, 0.8, 1.2, 2.0, 3.0\}$.[7]

Columns (1)–(5) in the first three rows of Table 2 report the estimated effects of the policy recommendation A_i. We normalize the estimates by dividing the original numbers by the sample outcome means for confidentiality reasons. The results show that the effects of the policy recommendation A_i on the purchase value, the number of transactions, and point usage are 35–92%, 43–74%, and 37–71% of their sample means, respectively. These positive effects mark a sharp

[7]Unlike the theoretical framework, the feature vector X_i consists of discrete and continuous variables. We compute APS by fixing the value of the discrete part and computing by simulation the APS integral with respect to the continuous part. See Appendix E for details.

Table 2: Off-policy evaluation using policy's generated data

	Our Proposed Method with APS Controls					Mean
	$\delta = 0.4$ (1)	$\delta = 0.8$ (2)	$\delta = 1.2$ (3)	$\delta = 2.0$ (4)	$\delta = 3.0$ (5)	Differences (6)
Effect on Purchase Value	0.35	0.82	0.92	0.54	0.72	-0.17
	(0.59)	(0.39)	(0.30)	(0.28)	(0.21)	(0.11)
Effect on # of Transactions	0.43	0.47	0.66	0.49	0.74	-0.07
	(0.50)	(0.34)	(0.28)	(0.25)	(0.19)	(0.10)
Effect on Point Usage	0.37	0.71	0.57	0.47	0.64	0.68
	(0.42)	(0.29)	(0.26)	(0.22)	(0.17)	(0.04)
Coupon Cost Effectiveness Measure	79.57	96.35	134	93.51	92.07	—
	(130)	(48.97)	(61.97)	(49.33)	(28.45)	
N	2758	4688	6016	8085	9602	89486

Notes: The first three rows of this table report estimated effects of the policy recommendation A_i on purchase behavior. Columns (1)–(5) report estimates from our method with several choices of δ used to compute APS. Column (6) reports the outcome mean differences between those with $A_i = 1$ and $A_i = 0$. Each APS is computed by averaging 100 simulation draws of the logging policy's binary decision. All numbers in the first three rows are normalized by dividing the original estimates by the sample outcome means. The fourth row reports our measure of coupon cost effectiveness, which predicts how much the purchase value would increase in USD if we increased the cost of the campaign by 1 USD. Heteroskedasticity-robust standard errors are reported in parentheses. The last row reports the number of observations with nonzero APS for every action (Columns (1)–(5)) or the total sample size (Column (6)).

contrast with Column (6), which reports the simple differences in the outcome means between those with $A_i = 1$ and those with $A_i = 0$. The simple mean differences on the purchase value and the number of transactions are negative. These negative estimates suggest that the logging policy tends to recommend a coupon to the customers who have a low propensity to make purchases. Our proposed method corrects for this negative selection bias by controlling for APS.

Evaluation of Counterfactual Policies.

The company needs to compensate for the discount that customers get by using points. Thus, adopting a new policy would be profitable only when the increase in revenue is sufficiently large compared to that in point usage. The company charges sellers 10% of every payment from the buyer, which means the revenue increases by 10% of the increase in purchase value. Hence, the policy change is beneficial if the ratio of the increases in the average purchase value and point usage is larger than 10.

Suppose we change our policy from ML to a counterfactual one π. Let Y_i^1 and Y_i^2 denote the purchase value and point usage respectively. Under the constant conditional effect assumption, i.e., $E[Y_i^1(1) - Y_i^1(0)|X_i] =: \beta$ and $E[Y_i^2(1) - Y_i^2(0)|X_i] =: \gamma$, the ratio can be obtained as follows:

$$\frac{E[\sum_{a \in \mathcal{A}} Y_i^1(a)\pi(a|X_i)] - E[\sum_{a \in \mathcal{A}} Y_i^1(a)ML(a|X_i)]}{E[\sum_{a \in \mathcal{A}} Y_i^2(a)\pi(a|X_i)] - E[\sum_{a \in \mathcal{A}} Y_i^2(a)ML(a|X_i)]} = \frac{\beta E[\pi(1|X_i) - ML(1|X_i)]}{\gamma E[\pi(1|X_i) - ML(1|X_i)]} = \frac{\beta}{\gamma}.$$

Row 4 in Table 2 reports the estimates of the ratio β/γ. The estimates are larger than 10 for all δ's. This result suggests that it would be profitable to expand the campaign.

7 Conclusion

We develop an OPE method for a class of logging policies including deficient support ones. Our method is based on the newly developed "Approximate Propensity Score." We prove that our estimator is consistent and demonstrate its practical performance through simulations and a real-world application. Promising directions for future work include developing a data-driven procedure to optimize the bandwidth. Also, the assumption of constant conditional mean reward differences may not be plausible in some applications. It will be challenging but interesting to relax this assumption to allow for certain types of heterogeneity. Finally, we look forward to applications of our method in a variety of business, policy, and scientific domains using machine learning.

References

Angrist, J. D.; and Pischke, J.-S. 2008. *Mostly harmless econometrics: an empiricist's companion*. Princeton University Press.

Beygelzimer, A.; and Langford, J. 2009. The offset tree for learning with partial labels. *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 129–138.

Crasta, G.; and Malusa, A. 2007. The Distance Function from the Boundary in a Minkowski Space. *Transactions of the American Mathematical Society*, 359: 5725–5759.

Duan, Y.; Jia, Z.; and Wang, M. 2020. Minimax-optimal off-policy evaluation with linear function approximation. *Proceedings of the 37th International Conference on Machine Learning*, 2701–2709.

Dudík, M.; Erhan, D.; Langford, J.; and Li, L. 2014. Doubly robust policy evaluation and optimization. *Statistical Science*, 29(4): 485–511.

Farajtabar, M.; Chow, Y.; and Ghavamzadeh, M. 2018. More robust doubly robust off-policy evaluation. *Proceedings of the 35th International Conference on Machine Learning*, 80: 1447–1456.

Hull, P. 2018. Subtracting the propensity score in linear models. *Working Paper*.

Krantz, S. G.; and Parks, H. R. 2008. *Geometric Integration Theory*. Birkhäuser Basel.

Kuzborskij, I.; Vernade, C.; Gyorgy, A.; and Szepesvari, C. 2021. Confident Off-Policy Evaluation and Selection through Self-Normalized Importance Weighting. *Proceedings of The 24th International Conference on Artificial Intelligence and Statistics*, 640–648.

Lee, D. S.; and Lemieux, T. 2010. Regression discontinuity designs in economics. *Journal of Economic Literature*, 48(2).

Li, S. 2011. Concise Formulas for the Area and Volume of a Hyperspherical Cap. *Asian Journal of Mathematics and Statistics*, 4: 66–70.

Precup, D. 2000. Eligibility traces for off-policy policy evaluation. *Proceedings of the Seventeenth International Conference on Machine Learning*, 759–766.

Rosenbaum, P. R.; and Rubin, D. B. 1983. The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70(1): 41–55.

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

Strehl, A.; Langford, J.; Li, L.; and Kakade, S. M. 2010. Learning from logged implicit exploration data. *Proceedings of the 23rd International Conference on Neural Information Processing Systems*, 2217–2225.

Su, Y.; Dimakopoulou, M.; Krishnamurthy, A.; and Dudik, M. 2020. Doubly robust off-policy evaluation with shrinkage. *Proceedings of the 37th International Conference on Machine Learning*, 119: 9167–9176.

Swaminathan, A.; and Joachims, T. 2015. The self-normalized estimator for counterfactual learning. *Proceedings of the 28th International Conference on Neural Information Processing Systems*, 3231–3239.

Wager, S.; and Athey, S. 2018. Estimation and inference of heterogeneous treatment effects using random forests. *Journal of the American Statistical Association*, 113(523): 1228–1242.

A Definitions

Definition A.1 (Twice continuously differentiable)**.** We say that a bounded open set $S \subset \mathbb{R}^p$ is *twice continuously differentiable* if for every $x \in S$, there exists a ball $B(x,\epsilon)$ and a one-to-one mapping ψ from $B(x,\epsilon)$ onto an open set $D \subset \mathbb{R}^p$ such that ψ and ψ^{-1} are twice continuously differentiable, $\psi(B(x,\epsilon)\cap S) \subset \{(x_1,...,x_p) \in \mathbb{R}^p : x_p > 0\}$ and $\psi(B(x,\epsilon)\cap \partial S) \subset \{(x_1,...,x_p) \in \mathbb{R}^p : x_p = 0\}$, where ∂S is the boundary of S.

Definition A.2 (k-dimensional Hausdorff measure)**.** The k-dimensional Hausdorff measure on $\mathbb{R}^p$ is defined as follows. Let Σ be the Lebesgue σ-algebra on $\mathbb{R}^p$ (the set of all Lebesgue measurable sets on $\mathbb{R}^p$). For $S \in \Sigma$ and $\delta > 0$, let $\mathcal{H}^k_\delta(S) = \inf\{\sum_{j=1}^\infty d(E_j)^k : S \subset \cup_{j=1}^\infty E_j, d(E_j) < \delta, E_j \subset \mathbb{R}^p \text{ for all } j\}$, where $d(E) = \sup\{\|x-y\| : x, y \in E\}$. The k-dimensional Hausdorff measure of A on $\mathbb{R}^p$ is $\mathcal{H}^k(S) = \lim_{\delta\to 0}\mathcal{H}^k_\delta(S)$.

B Discussion about Assumption 3

Here we discuss (1) what would happen if we drop Assumption 3 and (2) a potential way of relaxing this.

1. As explained in the last paragraph of Section 4, without this assumption, $\hat{\beta}_a$ is a consistent estimator of the mean reward difference for the subpopulation on the decision boundary between actions a and 1. The estimates may still allow us to derive a meaningful policy implication. For example, let us focus on the binary-action case, i.e., $:= \{0,1\}$. (Our application in Section 6 falls into this category; $A_i = 1$ means giving a coupon to user i.) Although we cannot estimate the conditional average effect $E[Y(1) - Y(0)|X = x]$ for each possible vector of the user characteristics $x \in \mathcal{X}$, we can still estimate the *local average effect*. Suppose that the logging policy ML is a threshold policy such that $ML(1 \mid x) = 1$ iff $\tau(x) \geq c$ for some score function τ and threshold c. (This is the case in our application. We also assume $\tau(X)$ is continuous.) Then, we can estimate the effect of coupon distribution for users near the decision boundary, i.e., $\Pi := E[Y(1) - Y(0) \mid \tau(X) = c]$. By giving an extra coupon to a customer whose score is slightly below the threshold c, the firm can increase its profit by Π.
2. One way to relax this assumption is to consider a partition of $\mathcal{X}$ and assume that the conditional mean difference between any two actions is constant within each cell in the partition. This allows the conditional mean differences to vary across cells. If for each (a, a') pair, each cell contains x such that $p^{ML}(a|x) > 0$ and $p^{ML}(a'|x) > 0$, we can consistently estimate the conditional mean differences and the expected reward from any policy. How to find such a partition is an interesting future topic.

C Comprehensive Discussion about Assumption 6

Assumption 6 (a)–(e) are a set of conditions we require for proving consistency of $\hat{\beta}_a$ when $\widetilde{ML}(1|x) > 0$ and $\widetilde{ML}(a|x) > 0$ for some $x \in \mathcal{X}$. Assumption 6 (b) allows the function ML to be discontinuous on a set of points with the Lebesgue measure zero. For example, ML is allowed to be a discontinuous step function as long as it is continuous almost everywhere. Assumption 6 (c) holds if the Lebesgue measures of the boundaries of $\mathcal{X}^a_{a,1}$ and $\mathcal{X}^1_{a,1}$ are zero.

Assumption 6 (e) rules out potential multicollinearity. If the support of $\widetilde{ML}(a|X_i)$ contains only one value in $(0,1)$, $q^{ML}_{\delta_n}(a|X_i)$ is asymptotically constant and equal to $\widetilde{ML}(a|X_i)$ conditional on $q^{ML}_{\delta_n}(a|X_i) \in (0,1)$, resulting in multicollinearity between $q^{ML}_{\delta_n}(a|X_i)$ and the intercept. Although dropping the intercept from the linear regression (2) solves this issue, Assumption 6 (e) allows us to only consider the regression with a intercept for the purpose of simplyfing the presentation.

Assumption 6 (f)–(i) are a set of additional conditions we require for proving consistency of $\hat{\beta}_a$ when $\widetilde{ML}(a|x)$ is either 0 or 1. In particular, we assume by Part (f) that the original logging policy ML is deterministic and the context space is partitioned into m groups based on the action that the logging policy chooses. $\partial\Omega^*_a$ then corresponds to the decision boundary for action a. In this case, the subsample for which $q^{ML}_{\delta_n}(a|X_i) \in (0,1)$ is contained by the δ_n-neighborhood of $\partial\Omega^*_a$.

Assumption 6 (g) imposes the differentiability of Ω^*_a. The conditions are satisfied if, for example, $\Omega^*_a = \{x \in \mathbb{R}^p : f(x) \geq 0\}$ for some twice continuously differentiable function $f : \mathbb{R}^p \to \mathbb{R}$ such that the gradient $\nabla f(x)$ is nonzero for all $x \in \mathbb{R}^p$ with $f(x) = 0$. In general, the differentiability of Ω^*_a may not hold. For example, if tree-based algorithms are used to partition the context space, the decision boundary $\partial\Omega^*_a$ is not differentiable. However, Assumption 6 (g) approximately holds in that Ω^*_a is arbitrarily well approximated by a set that satisfies the differentiability condition.

Part (1) of Assumption 6 (h) says that $\partial\Omega^*_a$ is $(p-1)$ dimensional and has nonzero density. Part (2) requires that the logging policy chooses either action 1 or a near the boundary of Ω^*_a even if the context value is not in the subsample $\mathcal{X}_{a,1}$ as long as it is in the neighborhood of $\mathcal{X}_{a,1}$.

Lastly, Assumption 6 (i) imposes continuity and boundedness on the conditional moments of rewards and the probability density near the boundary of Ω^*_a.

D Simulation Experiments: Details and Additional Results

Implementation Details

Parameter Choice. For the variance-covariance matrix $\boldsymbol{\Sigma}$ of X_i, we first create a 100×100 symmetric matrix $\boldsymbol{V}$ such that the diagonal elements are one, $\boldsymbol{V}_{ij}$ is nonzero and equal to $\boldsymbol{V}_{ji}$ for $(i,j) \in \{2,3,4,5,6\} \times \{35,66,78\}$, and everything else is zero. We draw values from Unif$(-0.5, 0.5)$ independently for the nonzero off-diagonal elements of $\boldsymbol{V}$. We then create matrix $\boldsymbol{\Sigma} = \boldsymbol{V} \times \boldsymbol{V}$, which is positive semidefinite.

For α_0 and α_a, we first draw $\tilde{\alpha}_{0,j}$, $j = 51, \ldots, 100$ from Unif$(-100, 100)$ independently across j, and draw $\tilde{\alpha}_{a,j}$, $j = 1, \ldots, 100$ from Unif$(-150, 200)$ independently across j and actions a. We then set $\tilde{\alpha}_{0,j} = \frac{1}{5}\sum_{a=1}^{5} \tilde{a}_{a,j}$ for $j = 1, \ldots, 50$ and calculate α_0 and α_a by normalizing $\tilde{\alpha}_0$ and $\tilde{\alpha}_a$ such that $\mathrm{Var}(\sum_{k=1}^{100} X_{ki}\alpha_{0,k}) = \mathrm{Var}(\sum_{k=1}^{100} X_{ki}\alpha_{a,k}) = 1$ for all actions a.

Independent Training Sample $\tilde{\mathcal{D}}$. Before simulating 1,000 hypothetical samples, we construct an independent sample $\tilde{\mathcal{D}} = \{(\tilde{Y}_i, \tilde{X}_i, \tilde{A}_i)\}_{i=1}^{\tilde{n}}$ of size $\tilde{n}$ = 10,000. The distribution of $(\tilde{Y}_i, \tilde{X}_i, \tilde{A}_i)$ is the same as that of (Y_i, X_i, A_i) except that (1) $\tilde{Y}_i(a)$ is generated by $\tilde{Y}_i(a) = \sum_{k=1}^{100} X_{ki}^2(0.75\alpha_{0,k} + 0.5\alpha_{a,k}) + 0.25u_i + 0.5\epsilon_i(a)$, where $\epsilon_i(a) \sim N(0,1)$ and $\alpha_a = (\alpha_{a,1}, ..., \alpha_{a,100}) \in \mathbb{R}^{100}$, and (2) $\Pr(\tilde{A}_i = a) = 1/5$ for all actions a. This can be viewed as data from a past A/B test conducted to construct a policy.

Construction of Reward Prediction Functions τ_{pred}^{ML} and τ_{pred}^{π}. We use $\tilde{\mathcal{D}}$ to fit a linear model $\tilde{Y}_i = \sum_{a=1}^{5}(b_a + \sum_{k=1}^{100} \tilde{X}_{ki}c_{a,k})1\left\{\tilde{A}_i = a\right\} + e_i$ and compute $\tau_{pred}^{ML}(x, a) = \hat{b}_a + \sum_{k=1}^{100} \tilde{x}_k\hat{c}_{a,k}$. We repeat this process using a new set of $\tilde{n}$ independent draws of $\tilde{A}_i$ to construct τ_{pred}^{π}. We construct τ_{pred}^{ML} and τ_{pred}^{π} only once, and use them for all of the 1,000 samples.

Training Upper Confident Bound. We use $\tilde{\mathcal{D}}$ to train an Upper Confidence Bound bandit algorithm as follows. Let $D_1(x_1) \in \{1, ..., 10\}$ indicate which decile of X_{1i} the individual with $\tilde{X}_{1i} = x_1$ belongs to. Define $D_2(x_2)$ analogously for $\tilde{X}_{2i}$. Let $Q(a, d_1, d_2)$ be the sample mean reward for each action a for every decile pair (d_1, d_2) in the distribution of $\tilde{X}_{1i}$ and $\tilde{X}_{2i}$. We then compute $UCB(x, a) = Q(a, D_1(x_1), D_2(x_2)) + c\sqrt{\frac{\log \tilde{n}}{\tilde{N}_{a,D_1(x_1),D_2(x_2)}}}$. Here, we set exploration parameter c to 2. $\tilde{n}(= 10,000)$ is the size of the training sample $\tilde{\mathcal{D}}$. $\tilde{N}_{a,d_1,d_2}$ is the number of observations with action a for the decile pair (d_1, d_2) in the sample.

Additional Results

We also consider the case in which the conditional mean reward differences are not constant over x: $Y_i(a)$ is generated as $Y_i(a) = \sum_{k=1}^{100} X_{ki}^2(0.75\alpha_{0,k} + \alpha_{a,k}) + 0.25u_i$, where $\alpha_a = (\alpha_{a,1}, ..., \alpha_{a,100}) \in \mathbb{R}^{100}$. The rest of the experiment setup is the same as that in Section 5.

Table D.0.1 reports the result. Our method does not necessarily outperform the alternatives, suggesting a limitation of our method when the conditional mean reward differences depend on the context.

E Approximate Propensity Score with Discrete Context Variables

In this section, we provide the definition of APS when X_i includes discrete context variables. Suppose that $X_i = (X_{di}, X_{ci})$, where $X_{di} \in \mathbb{R}^{p_d}$ is a vector of discrete context variables, and $X_{ci} \in \mathbb{R}^{p_c}$ is a vector of continuous context variables. Let $\mathcal{X}_d$ denote the support of X_{di} and be assumed to be finite. We also assume that X_{ci} is continuously distributed conditional on X_{di}.

We define APS as follows: for each $x = (x_d, x_c) \in \mathcal{X}$ and $a \in \mathcal{A}$,

$$p_\delta^{ML}(a|x) \equiv \frac{\int_{B(x_c,\delta)} ML(a|x_d, x_c^*)dx_c^*}{\int_{B(x_c,\delta)} dx_c^*},$$
$$p^{ML}(a|x) \equiv \lim_{\delta \to 0} p_\delta^{ML}(a|x),$$

where $B(x_c, \delta) = \{x_c^* \in \mathbb{R}^{p_c} : \|x_c - x_c^*\| \le \delta\}$ is the δ-ball around $x_c \in \mathbb{R}^{p_c}$. In other words, we take the average of the $ML(a|x_d, x_c^*)$ values when x_c^* is uniformly distributed on $B(x_c, \delta)$ holding x_d fixed, and let $\delta \to 0$.

F Notations and Lemmas

Basic Notations

For a vector or matrix X, we use X' to denote its transpose.

For a scalar-valued differentiable function $f : A \subset \mathbb{R}^n \to \mathbb{R}$, let $\nabla f : \mathbb{R}^n \to \mathbb{R}^n$ be a gradient of f: for every $x \in A$,

$$\nabla f(x) = \left(\frac{\partial f(x)}{\partial x_1}, \cdots, \frac{\partial f(x)}{\partial x_n}\right)'.$$

Also, when the second-order partial derivatives of f exist, let $D^2 f(x)$ be the Hessian matrix:

$$D^2 f(x) = \begin{bmatrix} \frac{\partial^2 f(x)}{\partial x_1^2} & \cdots & \frac{\partial^2 f(x)}{\partial x_1 \partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 f(x)}{\partial x_n \partial x_1} & \cdots & \frac{\partial^2 f(x)}{\partial x_n^2} \end{bmatrix}$$

Table D.0.1: Simulation results: non-constant conditional mean reward differences

	Our Proposed Method with APS Controls				Method with Mean Differences		Direct
	$\delta = 0.1$	$\delta = 0.5$	$\delta = 1$	$\delta = 2.5$	A/B Test Sample	Full Sample	Method
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Experiment 1: Mix of A/B Test and Deterministic Policy							
Bias	$-.012$	.015	.022	.018	$-.026$	$-.006$	—
S.D.	.045	.038	.033	.026	.045	.018	—
RMSE	.046	.041	.040	.031	.052	.019	—
Avg. N	1806	6009	11627	30136	500	50000	—
Experiment 2: Upper Confidence Bound							
Bias	$-.118$	$-.119$	$-.121$	$-.119$	—	—	$-.117$
S.D.	.027	.012	.009	.006	—	—	.006
RMSE	.121	.120	.121	.119	—	—	.117
Avg. N	3397	17343	31107	47601	—	—	50000

Notes: This table shows the bias, the standard deviation (S.D.), and the root mean squared error (RMSE) of the estimators of the reward from the counterfactual policy $V(\pi)$ in the two simulation experiments. We use $1,000$ simulations of a size $50,000$ sample to compute these statistics. Columns (1)–(4) report estimates from our method with several choices of δ. Each APS is computed by averaging 100 simulation draws of the ML value. In columns (5)–(6), we estimate the mean reward differences $\beta(a,1)$ by the sample mean differences in the A/B test segment and the full sample, respectively. In column (7), we estimate $\beta(a,1)$ by fitting a linear model that predicts the reward from the context and action. The bottom row of each panel shows the average number of observations with nonzero APS for every action (Columns (1)–(4)), that with nonzero ML for every action (Column (5)), or the total sample size (Columns (6)–(7)).

for each $x \in A$.

Let $f : A \subset \mathbb{R}^m \to \mathbb{R}^n$ be a function such that its first-order partial derivatives exist. For each $x \in A$, let $Jf(x)$ be the Jacobian matrix of f at x:

$$Jf(x) = \begin{bmatrix} \frac{\partial f_1(x)}{\partial x_1} & \cdots & \frac{\partial f_1(x)}{\partial x_m} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_n(x)}{\partial x_1} & \cdots & \frac{\partial f_n(x)}{\partial x_m} \end{bmatrix}.$$

For a positive integer n, let I_n denote the $n \times n$ identity matrix.

Differential Geometry

We provide some concepts and facts from differential geometry of twice continuously differentiable sets, following (Crasta and Malusa 2007). Let $A \subset \mathbb{R}^p$ be a twice continuously differentiable set. For each $x \in \partial A$, we denote by $\nu_A(x) \in \mathbb{R}^p$ the inward unit normal vector of ∂A at x, that is, the unit vector orthogonal to all vectors in the tangent space of ∂A at x that points toward the inside of A. For a set $A \subset \mathbb{R}^p$, let $d^s_A : \mathbb{R}^p \to \mathbb{R}$ be the signed distance function of A, defined by

$$d^s_A(x) = \begin{cases} d(x, \partial A) & \text{if } x \in \mathrm{cl}(A) \\ -d(x, \partial A) & \text{if } x \in \mathbb{R}^p \setminus \mathrm{cl}(A), \end{cases}$$

where $d(x,B) = \inf_{y \in B} \|y - x\|$ for any $x \in \mathbb{R}^p$ for a set $B \subset \mathbb{R}^p$. Note that we can write $N(\partial A, \delta) = \{x \in \mathbb{R}^p : -\delta < d^s_A(x) < \delta\}$ for $\delta > 0$. Lastly, let $\Pi_{\partial A}(x) = \{y \in \partial A : \|y - x\| = d(x, \partial A)\}$ be the set of projections of x on ∂A.

Lemma F.1 (Corollary of Theorem 4.16, (Crasta and Malusa 2007))**.** *Let $A \subset \mathbb{R}^p$ be nonempty, bounded, open, connected and twice continuously differentiable. Then the function d^s_A is twice continuously differentiable on $N(\partial A, \mu)$ for some $\mu > 0$. In addition, for every $x_0 \in \partial A$, $\Pi_{\partial A}(x_0 + t\nu_A(x_0)) = \{x_0\}$ for every $t \in (-\mu, \mu)$. Furthermore, for every $x \in N(\partial A, \mu)$, $\Pi_{\partial A}(x)$ is a singleton, $\nabla d^s_A(x) = \nu_A(y)$ and $x = y + d^s_A(x)\nu_A(y)$ for $y \in \Pi_{\partial A}(x)$, and $\|\nabla d^s_A(x)\| = 1$.*

Proof. We apply results from (Crasta and Malusa 2007). Let $K = \{x \in \mathbb{R}^p : \|x\| \leq 1\}$. K is nonempty, compact, convex subset of $\mathbb{R}^p$ with the origin as an interior point. The polar body of K, defined as $K_0 = \{y \in \mathbb{R}^p : y \cdot x \leq 1 \text{ for all } x \in K\}$, is

K itself. The gauge functions $\rho_K, \rho_{K_0} : \mathbb{R}^p \to [0, \infty]$ of K and K_0 are given by

$$\rho_K(x) \equiv \inf\{t \geq 0 : x \in tK\} = \|x\|,$$
$$\rho_{K_0}(x) \equiv \inf\{t \geq 0 : x \in tK_0\} = \|x\|.$$

Given ρ_{K_0}, the Minkowski distance from a set $S \subset \mathbb{R}^p$ is defined as

$$\delta_S(x) \equiv \inf_{y \in S} \rho_{K_0}(x - y), \quad x \in \mathbb{R}^p.$$

Note that we can write

$$d^s_A(x) = \begin{cases} \delta_{\partial A}(x) & \text{if } x \in \mathrm{cl}(A) \\ -\delta_{\partial A}(x) & \text{if } x \in \mathbb{R}^p \setminus \mathrm{cl}(A). \end{cases}$$

It then follows from Theorem 4.16 of (Crasta and Malusa 2007) that d^s_A is twice continuously differentiable on $N(\partial A, \mu)$ for some $\mu > 0$, and for every $x_0 \in \partial A$,

$$\begin{aligned} \nabla d^s_A(x_0) &= \frac{\nu_A(x_0)}{\rho_K(\nu_A(x_0))} \\ &= \frac{\nu_A(x_0)}{\|\nu_A(x_0)\|} \\ &= \nu_A(x_0), \end{aligned}$$

where the last equality follows since $\nu_A(x_0)$ is a unit vector. It then follows that $\|\nabla d^s_A(x_0)\| = \|\nu_A(x_0)\| = 1$ for every $x_0 \in \partial A$. Also, it is obvious that, for every $x_0 \in \partial A$, $\Pi_{\partial A}(x_0) = \{x_0\}$ and $x_0 = x_0 + d^s_A(x_0)\nu_A(x_0)$, since $d^s_A(x_0) = 0$. In addition, as stated in the proof of Theorem 4.16 of (Crasta and Malusa 2007), μ is chosen so that (4.7) in Proposition 4.6 of (Crasta and Malusa 2007) holds for every $x_0 \in \partial A$ and every $t \in (-\mu, \mu)$. That is, $\Pi_{\partial A}(x_0 + t\nabla\rho_K(\nu_A(x_0))) = \{x_0\}$ for every $x_0 \in \partial A$ and every $t \in (-\mu, \mu)$. Since $\nabla\rho_K(\nu_A(x_0)) = \frac{\nu_A(x_0)}{\|\nu_A(x_0)\|} = \nu_A(x_0)$, $\Pi_{\partial A}(x_0 + t\nu_A(x_0)) = \{x_0\}$ for every $x_0 \in \partial A$ and every $t \in (-\mu, \mu)$.

Furthermore, for every $x \in N(\partial A, \mu) \setminus \partial A$, $\Pi_{\partial A}(x)$ is a singleton as shown in the proof of Theorem 4.16 of (Crasta and Malusa 2007). Let $\pi_{\partial A}(x)$ be the unique element in $\Pi_{\partial A}(x)$. By Lemma 4.3 of (Crasta and Malusa 2007), for every $x \in N(\partial A, \mu) \setminus \partial A$,

$$\begin{aligned} \nabla d^s_A(x) &= \frac{\nu_A(\pi_{\partial A}(x))}{\rho_K(\nu_A(\pi_{\partial A}(x)))} \\ &= \frac{\nu_A(\pi_{\partial A}(x))}{\|\nu_A(\pi_{\partial A}(x))\|} \\ &= \nu_A(\pi_{\partial A}(x)), \end{aligned}$$

where the last equality follows since $\nu_A(\pi_{\partial A}(x))$ is a unit vector. It then follows that $\|\nabla d^s_A(x)\| = \|\nu_A(\pi_{\partial A}(x))\| = 1$ for every $x \in N(\partial A, \mu) \setminus \partial A$.

Lastly, note that

$$\delta_{\partial A}(x) = \begin{cases} d^s_A(x) & \text{if } x \in N(\partial A, \mu) \cap \mathrm{int}(A) \\ -d^s_A(x) & \text{if } x \in N(\partial A, \mu) \setminus \mathrm{cl}(A), \end{cases}$$

and

$$\nabla\delta_{\partial A}(x) = \begin{cases} \nabla d^s_A(x) & \text{if } x \in N(\partial A, \mu) \cap \mathrm{int}(A) \\ -\nabla d^s_A(x) & \text{if } x \in N(\partial A, \mu) \setminus \mathrm{cl}(A), \end{cases}$$

so $\delta_{\partial A}(x)\nabla\delta_{\partial A}(x) = d^s_A(x)\nabla d^s_A(x) = d^s_A(x)\nu_A(\pi_{\partial A}(x))$ for every $x \in N(\partial A, \mu) \setminus \partial A$. By Proposition 3.3 (i) of (Crasta and Malusa 2007), for every $x \in N(\partial A, \mu) \setminus \partial A$,

$$\nabla\rho_K(\nabla\delta_{\partial A}(x)) = \frac{x - \pi_{\partial A}(x)}{\delta_{\partial A}(x)},$$

which implies that

$$\begin{aligned} x &= \pi_{\partial A}(x) + \delta_{\partial A}(x)\nabla\rho_K(\nabla\delta_{\partial A}(x)) \\ &= \pi_{\partial A}(x) + \delta_{\partial A}(x)\frac{\nabla\delta_{\partial A}(x)}{\|\nabla\delta_{\partial A}(x)\|} \\ &= \pi_{\partial A}(x) + d^s_A(x)\nu_A(\pi_{\partial A}(x)). \end{aligned}$$

$\square$

We say that a set $A \subset \mathbb{R}^n$ is a *m-dimensional C^1 submanifold of* $\mathbb{R}^n$ if for every point $x \in A$, there exist an open neighborhood $V \subset \mathbb{R}^n$ of x and a one-to-one continuously differentiable function ϕ from an open set $U \subset \mathbb{R}^m$ to $\mathbb{R}^n$ such that the Jacobian matrix $J\phi(u)$ is of rank m for all $u \in U$, and $\phi(U) = V \cap A$.

Lemma F.2. *Let $A \subset \mathbb{R}^p$ be nonempty, bounded, open, connected and twice continuously differentiable. Then ∂A is a $(p-1)$-dimensional C^1 submanifold of $\mathbb{R}^p$,*

Proof. Fix any $x^* \in \partial A$. By Lemma F.1, $\nabla d^s_A(x^*)$ is nonzero. Without loss of generality, let $\frac{\partial d^s_A(x^*)}{\partial x_p} \neq 0$. Let $\psi : \mathbb{R}^p \to \mathbb{R}^p$ be the function such that $\psi(x) = (x_1, ..., x_{p-1}, d^s_A(x))$. ψ is continuously differentiable, and the Jacobian matrix of ψ at x^* is given by

$$J\psi(x^*) = \begin{pmatrix} \frac{\partial \psi_1}{\partial x_1}(x^*) & \cdots & \frac{\partial \psi_1}{\partial x_p}(x^*) \\ \vdots & \ddots & \vdots \\ \frac{\partial \psi_p}{\partial x_1}(x^*) & \cdots & \frac{\partial \psi_p}{\partial x_p}(x^*) \end{pmatrix} = \begin{pmatrix} & & & 0 \\ & I_{p-1} & & \vdots \\ & & & 0 \\ \frac{\partial d^s_A(x^*)}{\partial x_1} & \cdots & \frac{\partial d^s_A(x^*)}{\partial x_{p-1}} & \frac{\partial d^s_A(x^*)}{\partial x_p} \end{pmatrix}.$$

Since $\frac{\partial d^s_A(x^*)}{\partial x_p} \neq 0$, the Jacobian matrix is invertible. By the Inverse Function Theorem, there exist an open set V containing x^* and an open set W containing $\psi(x^*)$ such that $\psi : V \to W$ has an inverse function $\psi^{-1} : W \to V$ that is continuously differentiable. We make V small enough so that $\frac{\partial d^s_A(x)}{\partial x_p} \neq 0$ for every $x \in V$. The Jacobian matrix of ψ^{-1} is given by $J\psi^{-1}(y) = J\psi(\psi^{-1}(y))^{-1}$ for all $y \in W$.

Now note that $\psi(x) = (x_1, ..., x_{p-1}, 0)$ for all $x \in V \cap \partial A$ by the definition of d^s_A. Let $U = \{(x_1, ..., x_{p-1}) \in \mathbb{R}^{p-1} : x \in V \cap \partial A\}$ and $\phi : U \to \mathbb{R}^p$ be a function such that $\phi(u) = \psi^{-1}((u, 0))$ for all $u \in U$. Below we verify that ϕ is one-to-one and continosly differentiable, that $J\phi(u)$ is of rank $p-1$ for all $u \in U$, that $\phi(U) = V \cap \partial A$, and that U is open.

First, ϕ is one-to-one, since ψ^{-1} is one-to-one, and $(u, 0) \neq (u', 0)$ if $u \neq u'$. Second, ϕ is continuously differentiable, since ψ^{-1} is so. The Jacobian matrix of ϕ at $u \in U$ is by definition

$$J\phi(u) = \begin{pmatrix} \frac{\partial \psi_1^{-1}}{\partial y_1}((u,0)) & \cdots & \frac{\partial \psi_1^{-1}}{\partial y_{p-1}}((u,0)) \\ \vdots & \ddots & \vdots \\ \frac{\partial \psi_p^{'}-1}{\partial y_1}((u,0)) & \cdots & \frac{\partial \psi_p^{-1}}{\partial y_{p-1}}((u,0)) \end{pmatrix}.$$

Note that this is the left $p \times (p-1)$ submatrix of $J\psi^{-1}((u, 0))$. Since $J\psi^{-1}((u, 0))$ has full rank, $J\phi(u)$ is of rank $p-1$. Moreover,

$$\begin{aligned} \phi(U) &= \{\psi^{-1}((u, 0)) : u \in U\} \\ &= \{\psi^{-1}((x_1, ..., x_{p-1}, 0)) : x \in V \cap \partial A\} \\ &= \{\psi^{-1}(\psi(x)) : x \in V \cap \partial A\} \\ &= V \cap \partial A. \end{aligned}$$

Lastly, we show that U is open. Pick any $\bar{u} \in U$. Then, there exists $\bar{x}_p \in \mathbb{R}$ such that $(\bar{u}, \bar{x}_p) \in V \cap \partial A$. As $(\bar{u}, \bar{x}_p) \in V \cap \partial A$, $d^s_A((\bar{u}, \bar{x}_p)) = 0$. Since $\frac{\partial d^s_A((\bar{u}, \bar{x}_p))}{\partial x_p} \neq 0$, it follows by the Implicit Function Theorem that there exist an open set $S \subset \mathbb{R}^{p-1}$ containing $\bar{u}$ and a continuously differentiable function $g : S \to \mathbb{R}$ such that $g(\bar{u}) = \bar{x}_p$ and $d^s_A(u, g(u)) = 0$ for all $u \in S$. Since g is continuous, $(\bar{u}, g(\bar{u})) \in V$ and V is open, there exists an open set $S' \subset S$ containing $\bar{u}$ such that $(u, g(u)) \in V$ for all $u \in S'$. By the definition of d^s_A, $d^s_A(x) = 0$ if and only if $x \in \partial A$. Therefore, if $u \in S'$, $(u, g(u))$ must be contained by ∂A, for otherwise $d^s_A(u, g(u)) \neq 0$, which is a contradiction. Thus, $(u, g(u)) \in V \cap \partial A$ and hence $u \in U$ for all $u \in S'$. This implies that S' is an open subset of U containing $\bar{u}$, which proves that U is open. $\square$

Geometric Measure Theory

We provide some concepts and facts from geometric measure theory, following (Krantz and Parks 2008). Recall that for a function $f : A \subset \mathbb{R}^m \to \mathbb{R}^n$ and a point $x \in A$ at which f is differentiable, $Jf(x)$ denotes the Jacobian matrix of f at x.

Lemma F.3 (Coarea Formula, Lemma 5.1.4 and Corollary 5.2.6 of (Krantz and Parks 2008))**.** *If $f : \mathbb{R}^m \to \mathbb{R}^n$ is a Lipschitz function and $m \geq n$, then*

$$\int_A g(x) J_n f(x) d\mathcal{L}^m(x) = \int_{\mathbb{R}^n} \int_{\{x' \in A : f(x') = y\}} g(x) d\mathcal{H}^{m-n}(x) d\mathcal{L}^n(y)$$

for every Lebesgue measurable subset A of $\mathbb{R}^m$ and every $\mathcal{L}^m$-measurable function $g : A \to \mathbb{R}$, where for each $x \in \mathbb{R}^m$ at which f is differentiable,

$$J_n f(x) = \sqrt{\det((Jf(x))(Jf(x))')}.$$

Let A be an m-dimensional C^1 submanifold of $\mathbb{R}^n$. Let $x \in A$ and let $\phi : U \subset \mathbb{R}^m \to \mathbb{R}^n$ be as in the definition of m-dimensional C^1 submanifold. We denote by $T_A(x)$ the tangent space of A at x, $\{J\phi(u)v : v \in \mathbb{R}^m\}$, where $u = \phi^{-1}(x)$.

Lemma F.4 (Area Formula, Lemma 5.3.5 and Theorem 5.3.7 of (Krantz and Parks 2008))**.** *Suppose $m \leq \nu$ and $f : \mathbb{R}^n \to \mathbb{R}^\nu$ is Lipschitz. If A is an m-dimensional C^1 submanifold of $\mathbb{R}^n$, then*

$$\int_A g(x) J_m^A f(x) d\mathcal{H}^m(x) = \int_{\mathbb{R}^\nu} \sum_{x \in A : f(x) = y} g(x) d\mathcal{H}^m(y)$$

for every $\mathcal{H}^m$-measurable function $g : A \to \mathbb{R}$, where for each $x \in \mathbb{R}^n$ at which f is differentiable,

$$J_m^A f(x) = \frac{\mathcal{H}^m(\{Jf(x)y : y \in P\})}{\mathcal{H}^m(P)}$$

for an arbitrary m-dimensional parallelepiped P contained in $T_A(x)$.

Let $A \subset \mathbb{R}^p$. For each $x \in \mathbb{R}^p$ at which d_A^s is differentiable and for each $\lambda \in \mathbb{R}$, let $\psi_A(x, \lambda) = x + \lambda \nabla d_A^s(x)$.

Lemma F.5. *Let $\Omega \subset \mathbb{R}^p$, and suppose that there exists a partition $\{\Omega_1, ..., \Omega_M\}$ of Ω such that*

(i) $\operatorname{dist}(\Omega_m, \Omega_{m'}) > 0$ for any $m, m' \in \{1, ..., M\}$ such that $m \neq m'$;

(ii) Ω_m is nonempty, bounded, open, connected and twice continuously differentiable for each $m \in \{1, ..., M\}$.

Then there exists $\mu > 0$ such that d_Ω^s is twice continuously differentiable on $N(\partial\Omega, \mu)$ and that

$$\int_{N(\partial\Omega, \delta)} g(x) dx = \int_{-\delta}^{\delta} \int_{\partial\Omega} g(u + \lambda \nu_\Omega(u)) J_{p-1}^{\partial\Omega} \psi_\Omega(u, \lambda) d\mathcal{H}^{p-1}(u) d\lambda$$

for every $\delta \in (0, \mu)$ and every function $g : \mathbb{R}^p \to \mathbb{R}$ that is integrable on $N(\partial\Omega, \delta)$, where for each fixed $\lambda \in (-\mu, \mu)$, $J_{p-1}^{\partial\Omega} \psi_\Omega(\cdot, \lambda)$ is calculated by applying the operation $J_{p-1}^{\partial\Omega}$ to the function $\psi_\Omega(\cdot, \lambda)$. Futhermore, $J_{p-1}^{\partial\Omega} \psi_\Omega(x, \cdot)$ is continuously differentiable in λ and $J_{p-1}^{\partial\Omega} \psi_\Omega(x, 0) = 1$ for every $x \in \partial\Omega$, and $J_{p-1}^{\partial\Omega} \psi_\Omega(\cdot, \cdot)$ and $\frac{\partial J_{p-1}^{\partial\Omega} \psi_\Omega(\cdot, \cdot)}{\partial\lambda}$ are bounded on $\partial\Omega \times (-\mu, \mu)$.

Proof. Let $\bar{\mu} = \frac{1}{2} \min_{m, m' \in \{1, ..., M\}, m \neq m'} \operatorname{dist}(\Omega_m^*, \Omega_{m'})$ so that $\{N(\partial\Omega_m, \bar{\mu})\}_{m=1}^M$ is a partition of $N(\partial\Omega, \bar{\mu})$. Note that for every $m \in \{1, ..., M\}$, $d_\Omega^s(x) = d_{\Omega_m}^s(x)$ for every $x \in N(\partial\Omega_m, \bar{\mu})$. By Lemma F.1, for every $m \in \{1, ..., M\}$, there exists $\bar{\mu}_m > 0$ such that $d_{\Omega_m}^s$ is twice continuously differentiable on $N(\partial\Omega_m, \bar{\mu}_m)$. Letting $\mu \in (0, \min\{\bar{\mu}, \bar{\mu}_1, ..., \bar{\mu}_M\})$, we have that d_Ω^s is twice continuously differentiable on $N(\partial\Omega, \mu)$. This implies that d_Ω^s is Lipschitz on $N(\partial\Omega, \mu)$. For every $\delta \in (0, \mu)$ and every function $g : \mathbb{R}^p \to \mathbb{R}$ that is integrable on $N(\partial\Omega, \delta)$,

$$\begin{aligned}
\int_{N(\partial\Omega, \delta)} g(x) dx &= \int_{\{x' \in \mathbb{R}^p : d_\Omega^s(x') \in (-\delta, \delta)\}} g(x) \sqrt{\det(\|\nabla d_\Omega^s(x)\|)} dx \\
&= \int_{\{x' \in \mathbb{R}^p : d_\Omega^s(x') \in (-\delta, \delta)\}} g(x) \sqrt{\det(\nabla d_\Omega^s(x)' \nabla d_\Omega^s(x))} dx \\
&= \int_{\{x' \in \mathbb{R}^p : d_\Omega^s(x') \in (-\delta, \delta)\}} g(x) \sqrt{\det((J d_\Omega^s(x))(J d_\Omega^s(x))')} dx \\
&= \int_{\mathbb{R}} \int_{\{x' \in \mathbb{R}^p : d_\Omega^s(x') \in (-\delta, \delta), d_\Omega^s(x') = \lambda\}} g(x) d\mathcal{H}^{p-1}(x) d\lambda \\
&= \int_{-\delta}^{\delta} \int_{\{x' \in \mathbb{R}^p : d_\Omega^s(x') = \lambda\}} g(x) d\mathcal{H}^{p-1}(x) d\lambda, \qquad (4)
\end{aligned}$$

where the first equality follows since $\|\nabla d_\Omega^s(x)\| = 1$ for every $x \in N(\partial\Omega, \delta)$ by Lemma F.1, the third equality follows from the definition of the Jacobian matrix, and the fourth equality follows from Lemma F.3.

Let $\Gamma(\lambda) = \{x \in \mathbb{R}^p : d_\Omega^s(x) = \lambda\}$ for each $\lambda \in (-\mu, \mu)$. Since ∇d_Ω^s is differentiable on $N(\partial\Omega, \mu)$, $\psi_\Omega(x, \lambda)$ is defined on $N(\partial\Omega, \mu) \times \mathbb{R}$. We show that $\{\psi_\Omega(x_0, \lambda) : x_0 \in \partial\Omega\} \subset \Gamma(\lambda)$ for every $\lambda \in (-\mu, \mu)$. By Lemma F.1, for every $x_0 \in \partial\Omega$, $\psi_\Omega(x_0, \lambda) = x_0 + \lambda \nu_\Omega(x_0)$ and

$$\begin{aligned}
\Pi_{\partial\Omega}(\psi_\Omega(x_0, \lambda)) &= \Pi_{\partial\Omega}(x_0 + \lambda \nu_\Omega(x_0)) \\
&= \{x_0\}.
\end{aligned}$$

Hence,

$$\begin{aligned}
d(\psi_\Omega(x_0, \lambda), \partial\Omega) &= \|\psi_\Omega(x_0, \lambda) - x_0\| \\
&= \|\lambda \nu_\Omega(x_0)\| \\
&= |\lambda|.
\end{aligned}$$

Since $\nu_\Omega(x_0)$ is an inward normal vector, $\psi_\Omega(x_0,\lambda) \in \mathrm{cl}(A)$ if $0 \le \lambda < \mu$, and $\psi_\Omega(x,\lambda_0) \in \mathbb{R}^p \setminus \mathrm{cl}(A)$ if $-\mu < \lambda < 0$. It follows that

$$\begin{aligned} d_A^s(\psi_\Omega(x_0,\lambda)) &= \begin{cases} |\lambda| & \text{if } 0 \le \lambda < \mu \\ -|\lambda| & \text{if } \mu < \lambda < 0 \end{cases} \\ &= \lambda, \end{aligned}$$

so $\{\psi_\Omega(x_0,\lambda) : x_0 \in \partial\Omega\} \subset \Gamma(\lambda)$. It also holds that $\Gamma(\lambda) \subset \{\psi_\Omega(x_0,\lambda) : x_0 \in \partial\Omega\}$, since by Lemma F.1, for every $x \in \Gamma(\lambda)$,

$$\begin{aligned} \psi_\Omega(\pi_{\partial\Omega}(x),\lambda) &= \pi_{\partial\Omega}(x) + \lambda \nabla d_\Omega^s(\pi_{\partial\Omega}(x)) \\ &= \pi_{\partial\Omega}(x) + d_\Omega^s(x)\nu_\Omega(\pi_{\partial\Omega}(x)) \\ &= x, \end{aligned}$$

where $\pi_{\partial\Omega}(x)$ is the unique element in $\Pi_{\partial\Omega}(x)$. Thus, $\{\psi_\Omega(x_0,\lambda) : x_0 \in \partial\Omega\} = \Gamma(\lambda)$.

Now note that $\{\partial\Omega_m\}_{m=1}^M$ is a partition of $\partial\Omega$, since $\mathrm{dist}(\Omega_m,\Omega_{m'}) > 0$ for any $m, m' \in \{1,...,M\}$ such that $m \neq m'$. By Lemma F.2, $\partial\Omega_m$ is a $(p-1)$-dimensional C^1 submanifold of $\mathbb{R}^p$ for every $m \in \{1,...,M\}$, and hence $\partial\Omega$ is a $(p-1)$-dimensional C^1 submanifold of $\mathbb{R}^p$. Furthermore, since ∇d_Ω^s is continuously differentiable on $N(\partial\Omega,\mu)$, $\psi_\Omega(\cdot,\lambda)$ is continuously differentiable on $N(\partial\Omega,\mu)$, which implies that $\psi_\Omega(\cdot,\lambda)$ is Lipschitz on $N(\partial\Omega,\mu)$ for every $\lambda \in \mathbb{R}$. Applying Lemma F.4, we have that for every $\lambda \in (-\mu,\mu)$,

$$\begin{aligned} \int_{\partial\Omega} g(u+\lambda\nu_\Omega(u)) J_{p-1}^{\partial\Omega}\psi_\Omega(u,\lambda) d\mathcal{H}^{p-1}(u) &= \int_{\partial\Omega} g(\psi_\Omega(u,\lambda)) J_{p-1}^{\partial\Omega}\psi_\Omega(u,\lambda) d\mathcal{H}^{p-1}(u) \\ &= \int_{\mathbb{R}^p} \sum_{u\in\partial\Omega:\psi_\Omega(u,\lambda)=x} g(\psi_\Omega(u,\lambda)) d\mathcal{H}^{p-1}(x). \end{aligned} \tag{5}$$

If $x \notin \{\psi_\Omega(u,\lambda) : u \in \partial\Omega\}$, $\{u \in \partial\Omega : \psi_\Omega(u,\lambda) = x\} = \emptyset$. If $x \in \{\psi_\Omega(u,\lambda) : u \in \partial\Omega\}$, there exists $u \in \partial\Omega$ such that $x = \psi_\Omega(u,\lambda)$. Since $\Pi_{\partial\Omega}(x) = \Pi_{\partial\Omega}(u+\lambda\nabla d_\Omega^s(u)) = \Pi_{\partial\Omega}(u+\lambda\nu_\Omega(u)) = \{u\}$ by Lemma F.1, such u is unique, and hence $\{u \in \partial\Omega : \psi_\Omega(u,\lambda) = x\}$ is a singleton. It follow that

$$\begin{aligned} \int_{\mathbb{R}^p} \sum_{u\in\partial\Omega:\psi_\Omega(u,\lambda)=x} g(\psi_\Omega(u,\lambda)) d\mathcal{H}^{p-1}(x) &= \int_{\{\psi_\Omega(u,\lambda):u\in\partial\Omega\}} g(x) d\mathcal{H}^{p-1}(x) \\ &= \int_{\Gamma(\lambda)} g(x) d\mathcal{H}^{p-1}(x), \end{aligned} \tag{6}$$

where the last equality holds since $\{\psi_\Omega(u,\lambda) : u \in \partial\Omega\} = \Gamma(\lambda)$. Combining (4), (5) and (6), we obtain

$$\int_{N(\partial\Omega,\delta)} g(x)dx = \int_{-\delta}^{\delta} \int_{\partial\Omega} g(u+\lambda\nu_\Omega(u)) J_{p-1}^{\partial\Omega}\psi_\Omega(u,\lambda) d\mathcal{H}^{p-1}(u) d\lambda.$$

We next show that $J_{p-1}^{\partial\Omega}\psi_\Omega(x,\cdot)$ is continuously differentiable in λ and $J_{p-1}^{\partial\Omega}\psi_\Omega(x,0) = 1$ for every $x \in \partial\Omega$. Fix an $x \in \partial\Omega$, and let $V_\Omega(x)$ be an arbitrary $p \times (p-1)$ matrix whose columns $v_1(x),...,v_{p-1}(x) \in \mathbb{R}^p$ form an orthonormal basis of $T_{\partial\Omega}(x)$. Let $P(x) \subset T_{\partial\Omega}(x)$ be a parallelepiped determined by $v_1(x),...,v_{p-1}(x)$, that is, let $P(x) = \{\sum_{k=1}^{p-1} c_k v_k(x) : 0 \le c_k \le 1 \text{ for } k = 1,...,p-1\}$. Since $v_1(x),...,v_{p-1}(x)$ are linearly independent, $P(x)$ is a $(p-1)$-dimensional parallelepiped. It follows that for each fixed $\lambda \in \mathbb{R}$,

$$\begin{aligned} \{J\psi_\Omega(x,\lambda)y : y \in P(x)\} &= \{J\psi_\Omega(x,\lambda)\sum_{k=1}^{p-1} c_k v_k(x) : 0 \le c_k \le 1 \text{ for } k = 1,...,p-1\} \\ &= \{\sum_{k=1}^{p-1} c_k J\psi_\Omega(x,\lambda) v_k(x) : 0 \le c_k \le 1 \text{ for } k = 1,...,p-1\} \\ &= \{\sum_{k=1}^{p-1} c_k w_k(x,\lambda) : 0 \le c_k \le 1 \text{ for } k = 1,...,p-1\}, \end{aligned}$$

where $w_k(x,\lambda) = J\psi_\Omega(x,\lambda)v_k(x)$ for $k = 1,...,p-1$. Since $J\psi_\Omega(x,\lambda)v_k(x)$ is the k-th column of $J\psi_\Omega(x,\lambda)V_\Omega(x)$, $\{J\psi_\Omega(x,\lambda)y : y \in P(x)\}$ is the parallelepiped determined by the columns of $J\psi_\Omega(x,\lambda)V_\Omega(x)$. By Proposition 5.1.2 of

(Krantz and Parks 2008), we have that

$$\begin{aligned}
J^{\partial\Omega}_{p-1}\psi_\Omega(x,\lambda) &= \frac{\mathcal{H}^{p-1}(\{\sum_{k=1}^{p-1} c_k w_k(x,\lambda) : 0 \le c_k \le 1 \text{ for } k=1,...,p-1\})}{\mathcal{H}^{p-1}(P(x))} \\
&= \frac{\sqrt{\det((J\psi_\Omega(x,\lambda)V_\Omega(x))'(J\psi_\Omega(x,\lambda)V_\Omega(x)))}}{\sqrt{\det(V_\Omega(x)'V_\Omega(x))}} \\
&= \frac{\sqrt{\det((V_\Omega(x)+\lambda D^2 d^s_\Omega(x)V_\Omega(x))'(V_\Omega(x)+\lambda D^2 d^s_\Omega(x)V_\Omega(x)))}}{\sqrt{\det(I_{p-1})}} \\
&= \sqrt{\det(V_\Omega(x)'V_\Omega(x)+2V_\Omega(x)'\lambda D^2 d^s_\Omega(x)V_\Omega(x)+V_\Omega(x)'(\lambda D^2 d^s_\Omega(x))^2 V_\Omega(x))} \\
&= \sqrt{\det(I_{p-1}+\lambda V_\Omega(x)'(2D^2 d^s_\Omega(x)+\lambda(D^2 d^s_\Omega(x))^2)V_\Omega(x)))} \\
&= \sqrt{\det(I_p+\lambda V_\Omega(x)V_\Omega(x)'(2D^2 d^s_\Omega(x)+\lambda(D^2 d^s_\Omega(x))^2))},
\end{aligned}$$

where we use the fact that $V_\Omega(x)'V_\Omega(x) = I_{p-1}$ and the fact that $\det(I_m + AB) = \det(I_n + BA)$ for an $m \times n$ matrix A and an $n \times m$ matrix B (the Weinstein-Aronszajn identity). For every $x \in \partial\Omega$, $J^{\partial\Omega}_{p-1}\psi_\Omega(x,\cdot)$ is continuously differentiable in λ, and $J^{\partial\Omega}_{p-1}\psi_\Omega(x,0) = \sqrt{\det(I_p)} = 1$.

Lastly, we show that $J^{\partial\Omega}_{p-1}\psi_\Omega(\cdot,\cdot)$ and $\frac{\partial J^{\partial\Omega}_{p-1}\psi_\Omega(\cdot,\cdot)}{\partial\lambda}$ are bounded on $\partial\Omega \times (-\mu,\mu)$. Let $f, h : \partial\Omega \times \mathbb{R}^{p\times(p-1)} \to \mathbb{R}^{p\times p}$ be functions such that

$$\begin{aligned}
f(x,A) &= 2AA'D^2 d^s_\Omega(x), \\
h(x,A) &= AA'(D^2 d^s_\Omega(x))^2.
\end{aligned}$$

Also, let $k : \partial\Omega \times \mathbb{R} \times \mathbb{R}^{p\times(p-1)} \to \mathbb{R}$ be a function such that

$$k(x,\lambda,A) = \sqrt{\det(I_p + \lambda f(x,A) + \lambda^2 h(x,A))}.$$

Observe that

$$J^{\partial\Omega}_{p-1}\psi_\Omega(x,\lambda) = k(x,\lambda,V_\Omega(x))$$

and that

$$\begin{aligned}
&\frac{\partial J^{\partial\Omega}_{p-1}\psi_\Omega(x,\lambda)}{\partial\lambda} \\
&= \frac{\partial k(x,\lambda,A)}{\partial\lambda}\Big|_{A=V_\Omega(x)} \\
&= \frac{1}{2k(x,\lambda,A)}\sum_{i,j}\frac{\partial\det(I_p+\lambda f(x,A)+\lambda^2 h(x,A))}{\partial b_{ij}}(f_{ij}(x,A)+2\lambda h_{ij}(x,A))\Big|_{A=V_\Omega(x)},
\end{aligned}$$

where $\frac{\partial\det(B)}{\partial b_{ij}}$ denotes the partial derivative of the function $\det : \mathbb{R}^{p\times p} \to \mathbb{R}$ with respect to the (i,j) entry of B.

Note that $k(\cdot,\cdot,\cdot)$ and $\frac{\partial k(\cdot,\cdot,\cdot)}{\partial\lambda}$ are continuous on $\partial\Omega \times \mathbb{R} \times \mathbb{R}^{p\times(p-1)}$ (except at the points for which $k(x,\lambda,A) = 0$), since $\det$ is infinitely differentiable, and f and h are continuous on $\partial\Omega \times \mathbb{R}^{p\times(p-1)}$. Let $S = \{(x,\lambda,A) \in \partial\Omega \times [-\mu,\mu] \times \mathbb{R}^{p\times(p-1)} : \|a_j\| = 1 \text{ for } k = 1,...,p-1\}$, where a_j denotes the jth column of A. Since $k(\cdot,\cdot,\cdot)$ and $\frac{\partial k(\cdot,\cdot,\cdot)}{\partial\lambda}$ are continuous and S is closed and bounded, $\bar{k} = \max_{(x,\lambda,A)\in S}|k(x,\lambda,A)|$ and $\bar{k}' = \max_{(x,\lambda,A)\in S}|\frac{\partial k(x,\lambda,A)}{\partial\lambda}|$ exist. Since $(x,\lambda,V_\Omega(x)) \in S$ for every $(x,\lambda) \in \partial\Omega \times (-\mu,\mu)$, it follows that $|J^{\partial\Omega}_{p-1}\psi_\Omega(x,\lambda)| \le \bar{k}$ and $|\frac{\partial J^{\partial\Omega}_{p-1}\psi_\Omega(x,\lambda)}{\partial\lambda}| \le \bar{k}'$ for every $(x,\lambda) \in \partial\Omega \times (-\mu,\mu)$. □

Other Lemmas

Lemma F.6. *Fix any $a \in \{2,...,m\}$. Let $\{V_i\}_{i=1}^\infty$ be i.i.d. random variables such that $E[V_i^2] < \infty$. If Assumption 6 (b) – (d) hold, then for $l \ge 0$ and $m = 0,1$,*

$$\begin{aligned}
&E[V_i q^{ML}_\delta(a|X_i)^l 1\{q^{ML}_\delta(a|X_i) \in (0,1)\}^m 1\{A_i \in \{1,a\}\}] \\
&\to E[V_i \widetilde{ML}(a|X_i)^l 1\{\widetilde{ML}(a|X_i) \in (0,1)\} 1\{A_i \in \{1,a\}\}]
\end{aligned}$$

as $\delta \to 0$. Moreover, if, in addition, $\delta_n \to 0$ as $n \to \infty$, then for $l \geq 0$,

$$\begin{aligned}&\frac{1}{n}\sum_{i=1}^{n} V_i q^{ML}_{\delta_n}(a|X_i)^l 1\{q^{ML}_{\delta_n}(a|X_i) \in (0,1)\}1\{A_i \in \{1,a\}\}\\ &\xrightarrow{p} E[V_i \widetilde{ML}(a|X_i)^l 1\{\widetilde{ML}(a|X_i) \in (0,1)\}1\{A_i \in \{1,a\}\}]\end{aligned}$$

as $n \to \infty$.

Proof. Note that

$$\begin{aligned}&E[\frac{1}{n}\sum_{i=1}^{n} V_i q^{ML}_{\delta_n}(a|X_i)^l 1\{q^{ML}_{\delta_n}(a|X_i) \in (0,1)\}1\{A_i \in \{1,a\}\}]\\ &= E[V_i q^{ML}_{\delta_n}(a|X_i)^l 1\{q^{ML}_{\delta_n}(a|X_i) \in (0,1)\}1\{A_i \in \{1,a\}\}].\end{aligned}$$

We show that

$$\begin{aligned}&E[V_i q^{ML}_{\delta}(a|X_i)^l 1\{q^{ML}_{\delta}(a|X_i) \in (0,1)\}^m 1\{A_i \in \{1,a\}\}]\\ &\to E[V_i \widetilde{ML}(a|X_i)^l 1\{\widetilde{ML}(a|X_i) \in (0,1)\}^m 1\{A_i \in \{1,a\}\}]\end{aligned}$$

for $l \geq 0$ and $m = 0, 1$ as $\delta \to 0$, and that

$$\mathrm{Var}(\frac{1}{n}\sum_{i=1}^{n} V_i q^{ML}_{\delta_n}(a|X_i)^l 1\{q^{ML}_{\delta_n}(a|X_i) \in (0,1)\}1\{A_i \in \{1,a\}\}) \to 0$$

for $l \geq 0$ as $n \to \infty$. For the first part, we have

$$\begin{aligned}&E[V_i q^{ML}_{\delta}(a|X_i)^l 1\{q^{ML}_{\delta}(a|X_i) \in (0,1)\}^m 1\{A_i \in \{1,a\}\}]\\ &= E[E[V_i|X_i, A_i] q^{ML}_{\delta}(a|X_i)^l 1\{q^{ML}_{\delta}(a|X_i) \in (0,1)\}^m 1\{A_i \in \{1,a\}\}]\\ &= E[\sum_{a' \in \{1,a\}} E[V_i|X_i, A_i = a'] q^{ML}_{\delta}(a|X_i)^l 1\{q^{ML}_{\delta}(a|X_i) \in (0,1)\}^m ML(a'|X_i)]\\ &= \int_{\mathcal{X}_{a,1}} g(x) q^{ML}_{\delta}(a|x)^l 1\{q^{ML}_{\delta}(a|x) \in (0,1)\}^m f_X(x)dx,\end{aligned}$$

where $g(x) = \sum_{a' \in \{1,a\}} E[V_i|X_i = x, A_i = a'] ML(a'|x)$.

Suppose $ML(a|\cdot)$ and $ML(1|\cdot)$ are continuous at x and $\widetilde{ML}(a|x) \in (0,1)$. Then, with change of variables $u = \frac{x^* - x}{\delta}$, for $a' \in \{1,a\}$,

$$\begin{aligned}p^{ML}_{\delta}(a'|x) &= \frac{\int_{B(x,\delta)} ML(a'|x^*)dx^*}{\int_{B(x,\delta)} dx^*}\\ &= \frac{\delta^p \int_{B(\mathbf{0},1)} ML(a'|x + \delta u)du}{\delta^p \int_{B(\mathbf{0},1)} du}\\ &\to \frac{\int_{B(\mathbf{0},1)} ML(a'|x)du}{\delta^p \int_{B(\mathbf{0},1)} du} = ML(a'|x)\end{aligned}$$

as $\delta \to 0$, where the convergence follows from the Dominated Convergence Theorem. It follows that $\lim_{\delta\to 0} q^{ML}_{\delta}(a|x) = \frac{ML(a|x)}{ML(a|x)+ML(1|x)} = \widetilde{ML}(a|x) \in (0,1)$, and hence $q^{ML}_{\delta}(a|x) \in (0,1)$ for sufficiently small $\delta > 0$. Therefore, $1\{q^{ML}_{\delta}(a|x) \in (0,1)\} \to 1 = 1\{\widetilde{ML}(a|x) \in (0,1)\}$ as $\delta \to 0$.

Suppose $x \in \mathrm{int}(\mathcal{X}^a_{a,1}) \cup \mathrm{int}(\mathcal{X}^1_{a,1})$. Then $B(x,\delta) \subset \mathcal{X}^a_{a,1}$ or $B(x,\delta) \subset \mathcal{X}^1_{a,1}$ for sufficiently small $\delta > 0$ by the fact that $\mathrm{int}(\mathcal{X}^a_{a,1})$ and $\mathrm{int}(\mathcal{X}^1_{a,1})$ are open. Note that if $\widetilde{ML}(a|x') = 1$, then $\widetilde{ML}(1|x') = 0$. Hence if $B(x,\delta) \subset \mathcal{X}^a_{a,1}$, $p^{ML}_{\delta}(1|x) = 0$ so $q^{ML}_{\delta}(a|x) = 1$. Likewise, if $B(x,\delta) \subset \mathcal{X}^1_{a,1}$, $p^{ML}_{\delta}(a|x) = 0$ so $q^{ML}_{\delta}(a|x) = 0$. It follows that $1\{q^{ML}_{\delta}(a|x) \in (0,1)\} \to 0 = 1\{\widetilde{ML}(a|x) \in (0,1)\}$ as $\delta \to 0$.

Since $ML(a|\cdot)$ and $ML(1|\cdot)$ are continuous at x for almost every $x \in \mathcal{X}_{a,1}$ by Assumption 6 (b), and either $\widetilde{ML}(a|x) \in (0,1)$ or $x \in \mathrm{int}(\mathcal{X}^a_{a,1}) \cup \mathrm{int}(\mathcal{X}^1_{a,1})$ for almost every $x \in \mathcal{X}_{a,1}$ by Assumption 6 (c), the above results imply that

$\lim_{\delta\to 0} q_\delta^{ML}(a|x) = \widetilde{ML}(a|x)$ and $\lim_{\delta\to 0} 1\{q_\delta^{ML}(a|x) \in (0,1)\} = 1\{\widetilde{ML}(a|x) \in (0,1)\}$ for almost every $x \in \mathcal{X}_{a,1}$. By the Dominated Convergence Theorem,

$$
\begin{aligned}
&E[V_i q_\delta^{ML}(a|X_i)^l 1\{q_\delta^{ML}(a|X_i) \in (0,1)\}^m 1\{A_i \in \{1,a\}\}] \\
&\to \int_{\mathcal{X}_{a,1}} g(x)\widetilde{ML}(a|x)^l 1\{\widetilde{ML}(a|x) \in (0,1)\}^m f_X(x)dx \\
&= E[V_i \widetilde{ML}(a|X_i)^l 1\{\widetilde{ML}(a|X_i) \in (0,1)\}^m 1\{A_i \in \{1,a\}\}]
\end{aligned}
$$

as $\delta \to 0$. As for variance,

$$
\begin{aligned}
&\mathrm{Var}(\frac{1}{n}\sum_{i=1}^n V_i q_{\delta_n}^{ML}(a|X_i)^l 1\{q_{\delta_n}^{ML}(a|X_i) \in (0,1)\} 1\{A_i \in \{1,a\}\}) \\
&\le \frac{1}{n} E[V_i^2 q_{\delta_n}^{ML}(a|X_i)^{2l}(1\{q_{\delta_n}^{ML}(a|X_i) \in (0,1)\} 1\{A_i \in \{1,a\}\})^2] \\
&\le \frac{1}{n} E[V_i^2] \\
&\to 0
\end{aligned}
$$

as $n \to \infty$. $\square$

G Proofs

Derivation of Equation (1)

$$
\begin{aligned}
V(\pi) &= V(ML) + E[\sum_{a\in\mathcal{A}} E[Y(a)|X](\pi(a|X) - ML(a|X))] \\
&= V(ML) + E[\sum_{a\in\mathcal{A}} (E[Y(a)|X] - E[Y(1)|X])(\pi(a|X) - ML(a|X))] \\
&\quad + E[E[Y(1)|X]\sum_{a\in\mathcal{A}} (\pi(a|X) - ML(a|X))] \\
&= V(ML) + E[\sum_{a\in\mathcal{A}} \beta(a,1)(\pi(a|X) - ML(a|X))] \\
&= V(ML) + E[\sum_{a=2}^m \beta(a,1)(\pi(a|X) - ML(a|X))],
\end{aligned}
$$

where we use Assumption 3 and the fact that $\sum_{a\in\mathcal{A}}(\pi(a|X) - ML(a|X)) = 0$ for the third equality. $\square$

Proof of Lemma 2

Suppose that Assumption 1 holds. Pick $a \in \mathcal{A}$ and $x \in \mathrm{int}(\mathcal{X})$ such that $p^{ML}(a|x) > 0$. If $ML(a|x) > 0$, $E[Y|X = x, A = a] = E[Y(a)|X = x]$, since A is independent of $Y(a)$ conditional on X. $E[Y(a)|X = x]$ is thus identified. Suppose $ML(a|x) = 0$. Since $x \in \mathrm{int}(\mathcal{X})$, $B(x,\delta) \subset \mathcal{X}$ for any sufficiently small $\delta > 0$. Moreover, since $p^{ML}(a|x) = \lim_{\delta\to 0} p_\delta^{ML}(a|x) > 0$, $p_\delta^{ML}(a|x) > 0$ for any sufficiently small $\delta > 0$. This implies that we can find a point $x_\delta \in B(x,\delta)(\subset \mathcal{X})$ such that $ML(a|x_\delta) > 0$ for any sufficiently small $\delta > 0$, for otherwise $p_\delta^{ML}(a|x) = 0$. Noting that $x_\delta \to x$ as $\delta \to 0$,

$$
\begin{aligned}
\lim_{\delta\to 0} E[Y|X = x_\delta, A = a] &= \lim_{\delta\to 0} E[Y(a)|X = x_\delta] \\
&= E[Y(a)|X = x],
\end{aligned}
$$

where the first equality follows from conditional independence and the second from Assumption 1. $\square$

Proof of Proposition 5

We show that $E[Y(a)|X = x]$ is identified for every (a, x) pair. Since $E[Y(a)|X = x]$ is identified for at least one $a \in \mathcal{A}$ for every $x \in \mathcal{X}$, and $E[Y(a')|X = x] = E[Y(a)|X = x] + \beta(a', a)$ by Assumption 3, it suffices to show that $\beta(a', a)$ is identified for every (a', a) pair. This is equivalent to proving that $\beta(a, 1)$ is identified for every $a \in \{2, ..., m\}$, since $\beta(a', a) = \beta(a', 1) - \beta(a, 1)$.

Take any $a \in \{2, ..., m\}$ and let $\{a_1, ..., a_L\}$ be the sequence that satisfies the condition in Assumption 4. Under Assumption 1, Lemma 2 implies that for every $l \in \{1, ..., L-1\}$, $E[Y(a_{l+1})|X = x]$ and $E[Y(a_l)|X = x]$ are identified for some $x \in \mathrm{int}(\mathcal{X})$. This implies that $\beta(a_{l+1}, a_l)$ is identified for every $l \in \{1, ..., L-1\}$ under Assumption 3. Since

$$\beta(a, 1) = \beta(a_L, a_{L-1}) + \beta(a_{L-1}, a_{L-2}) + \cdots + \beta(a_2, a_1),$$

$\beta(a, 1)$ is also identified. □

Proof of Theorem 7

Fix any $a \in \{2, ..., m\}$ and consider the regression from the subsample assigned to either action a or 1 throughout the proof. For notational simplicity, we omit the argument a from $\widetilde{ML}(a|x)$ and $q_\delta^{ML}(a|x)$ and denote them by $\widetilde{ML}(x)$ and $q_\delta^{ML}(x)$. Let $\mathbf{Z}_i = (1, 1\{A_i = a\}, q_{\delta_n}^{ML}(X_i))'$, and $I_i = 1\{q_{\delta_n}^{ML}(X_i) \in (0,1)\}$. Let

$$\hat{\beta} = \begin{pmatrix} \hat{\alpha}_a \\ \hat{\beta}_a \\ \hat{\gamma}_a \end{pmatrix} = (\sum_{i=1}^n \mathbf{Z}_i \mathbf{Z}_i' I_i 1\{A_i \in \{1, a\}\})^{-1} \sum_{i=1}^n \mathbf{Z}_i Y_i I_i 1\{A_i \in \{1, a\}\}.$$

Below, we prove that $\hat{\beta}_a$ converges in probability to $\beta(a, 1)$. The theorem then immediately follows. Also, the proof of Step G.0.0.4 shows that if Assumption 3 does not hold for a deterministic logging policy, $\hat{\beta}_a$ converges in probability to

$$\frac{\int_{\partial\Omega^* \cap \mathcal{X}_{a,1}} E[Y_i(a) - Y_i(1)|X_i = x] f_X(x) d\mathcal{H}^{p-1}(x)}{\int_{\partial\Omega^* \cap \mathcal{X}_{a,1}} f_X(x) d\mathcal{H}^{p-1}(x)},$$

which is the mean reward difference for the subpopulation on the decision boundary between a and 1.

We provide proofs separately for the two cases, the case in which $\Pr(\widetilde{ML}(X_i) \in (0,1)) > 0$ and the case in which $\Pr(\widetilde{ML}(X_i) \in (0,1)) = 0$.

Consistency of $\hat{\beta}_a$ When $\Pr(\widetilde{ML}(X_i) \in (0,1)|A_i \in \{1,a\}) > 0$ Let $\tilde{\mathbf{Z}}_i = (1, 1\{A_i = a\}, \widetilde{ML}(X_i))'$ and $I_i^{ML} = 1\{\widetilde{ML}(X_i) \in (0,1)\}$. By Lemma F.6,

$$\begin{aligned}
\hat{\beta} &= (\sum_{i=1}^n \mathbf{Z}_i \mathbf{Z}_i' I_i 1\{A_i \in \{1, a\}\})^{-1} \sum_{i=1}^n \mathbf{Z}_i Y_i I_i 1\{A_i \in \{1, a\}\} \\
&\xrightarrow{p} (E[\tilde{\mathbf{Z}}_i \tilde{\mathbf{Z}}_i' I_i^{ML} 1\{A_i \in \{1, a\}\}])^{-1} E[\tilde{\mathbf{Z}}_i Y_i I_i^{ML} 1\{A_i \in \{1, a\}\}] \\
&= (E[\tilde{\mathbf{Z}}_i \tilde{\mathbf{Z}}_i' I_i^{ML} | A_i \in \{1, a\}])^{-1} E[\tilde{\mathbf{Z}}_i Y_i I_i^{ML} | A_i \in \{1, a\}]
\end{aligned}$$

provided that $E[\tilde{\mathbf{Z}}_i \tilde{\mathbf{Z}}_i' I_i^{ML} | A_i \in \{1, a\}]$ is invertible. After a few lines of algebra, we have

$$\begin{aligned}
&\det(E[\tilde{\mathbf{Z}}_i \tilde{\mathbf{Z}}_i' I_i^{ML} | A_i \in \{1, a\}]) \\
=& \Pr(I_i^{ML} = 1 | A_i \in \{1, a\})^2 \mathrm{Var}(\widetilde{ML}(X_i) | I_i^{ML} = 1, A_i \in \{1, a\}) \\
&\times E[\widetilde{ML}(X_i)(1 - \widetilde{ML}(X_i)) I_i^{ML} | A_i \in \{1, a\}] \\
=& \Pr(I_i^{ML} = 1 | A_i \in \{1, a\})^2 \mathrm{Var}(\widetilde{ML}(X_i) | I_i^{ML} = 1, A_i \in \{1, a\}) \\
&\times E[\widetilde{ML}(X_i)(1 - \widetilde{ML}(X_i)) | A_i \in \{1, a\}].
\end{aligned}$$

Therefore, $E[\tilde{\mathbf{Z}}_i \tilde{\mathbf{Z}}_i' I_i^{ML} | A_i \in \{1, a\}]$ is invertible, since $\Pr(I_i^{ML} = 1 | A_i \in \{1, a\}) > 0$, and $\mathrm{Var}(\widetilde{ML}(X_i) | I_i^{ML} = 1, A_i \in \{1, a\}) > 0$ under Assumption 6 (e).

Another few lines of algebra gives

$$(E[\tilde{\mathbf{Z}}_i \tilde{\mathbf{Z}}_i' I_i^{ML} | A_i \in \{1, a\}])^{-1} = \frac{1}{E[\widetilde{ML}(X_i)(1 - \widetilde{ML}(X_i)) | A_i \in \{1, a\}]} \begin{bmatrix} * & * & * \\ 0 & 1 & -1 \\ * & * & * \end{bmatrix}.$$

Observe that

$$\begin{aligned}
E[1\{A_i = a\} Y_i(a) | X_i, A_i \in \{1, a\}] &= E[1\{A_i = a\} | X_i, A_i \in \{1, a\}] E[Y_i(a) | X_i] \\
&= \frac{\Pr(A_i = a | X_i)}{\Pr(A_i \in \{1, a\} | X_i)} E[Y_i(a) | X_i] \\
&= \frac{ML(a|X_i)}{ML(a|X_i) + ML(1|X_i)} E[Y_i(a) | X_i] \\
&= \widetilde{ML}(X_i) E[Y_i(a) | X_i],
\end{aligned}$$

where the first equality follows from the assumption that A_i is independent of $Y_i(\cdot)$ conditional on X_i. Likewise,

$$E[1\{A_i = 1\}Y_i(1)|X_i, A_i \in \{1, a\}] = (1 - \widetilde{ML}(X_i))E[Y_i(1)|X_i].$$

Therefore,

$$\begin{aligned}
\hat{\beta}_a &\xrightarrow{p} \frac{E[1\{A_i = a\}Y_i I_i^{ML}|A_i \in \{1, a\}] - E[\widetilde{ML}(X_i)Y_i I_i^{ML}|A_i \in \{1, a\}]}{E[\widetilde{ML}(X_i)(1 - \widetilde{ML}(X_i))|A_i \in \{1, a\}]} \\
&= \frac{E[1\{A_i = a\}Y_i(a)I_i^{ML} - \widetilde{ML}(X_i)(1\{A_i = a\}Y_i(a) + 1\{A_i = 1\}Y_i(1))I_i^{ML}|A_i \in \{1, a\}]}{E[\widetilde{ML}(X_i)(1 - \widetilde{ML}(X_i))|A_i \in \{1, a\}]} \\
&= \frac{E[\widetilde{ML}(X_i)E[Y_i(a)|X_i]I_i^{ML} - \widetilde{ML}(X_i)(\widetilde{ML}(X_i)E[Y_i(a)|X_i]|A_i \in \{1, a\}]}{E[\widetilde{ML}(X_i)(1 - \widetilde{ML}(X_i)|A_i \in \{1, a\}]} \\
&\quad + \frac{E[(1 - \widetilde{ML}(X_i))E[Y_i(1)|X_i])I_i^{ML}|A_i \in \{1, a\}]}{E[\widetilde{ML}(X_i)(1 - \widetilde{ML}(X_i)|A_i \in \{1, a\}]} \\
&= \frac{E[\widetilde{ML}(X_i)(1 - \widetilde{ML}(X_i))E[Y_i(a) - Y_i(1)|X_i]I_i^{ML}|A_i \in \{1, a\}]}{E[\widetilde{ML}(X_i)(1 - \widetilde{ML}(X_i)|A_i \in \{1, a\}]} \\
&= \beta(a, 1).
\end{aligned}$$

□

Consistency of $\hat{\beta}_a$ When $\Pr(\widetilde{ML}(X_i) \in (0, 1)|A_i \in \{1, a\}) = 0$ For notational simplicity, we omit subscript a from Ω_a^* and denote it by Ω^*. We use the notation and results provided in Appendix F. By Lemma F.5, under Assumption 6 (g), there exists $\mu > 0$ such that $d_{\Omega^*}^s$ is twice continuously differentiable on $N(\partial\Omega^*, \mu)$ and that

$$\int_{N(\partial\Omega^*,\delta)} g(x)dx = \int_{-\delta}^{\delta}\int_{\partial\Omega^*} g(u + \lambda\nu_{\Omega^*}(u))J_{p-1}^{\partial\Omega^*}\psi_{\Omega^*}(u, \lambda)d\mathcal{H}^{p-1}(u)d\lambda$$

for every $\delta \in (0, \mu)$ and every function $g : \mathbb{R}^p \to \mathbb{R}$ that is integrable on $N(\partial\Omega^*, \delta)$.

Our proof proceeds in five steps.

Step G.0.0.1. For every $(u, v) \in \partial\Omega^* \cap N(\mathcal{X}_{a,1}, \bar{\delta}) \times (-1, 1)$, $\lim_{\delta\to 0} q_\delta^{ML}(u + \delta v\nu_{\Omega^*}(u)) = k(v)$, where

$$k(v) = \begin{cases} 1 - \frac{1}{2}I_{(1-v^2)}(\frac{p+1}{2}, \frac{1}{2}) & \text{for } v \in [0, 1) \\ \frac{1}{2}I_{(1-v^2)}(\frac{p+1}{2}, \frac{1}{2}) & \text{for } v \in (-1, 0). \end{cases}$$

Here $I_x(\alpha, \beta)$ is the regularized incomplete beta function (the cumulative distribution function of the beta distribution with shape parameters α and β).

Proof. By Assumption 6 (h) (2), there exists $\bar{\delta} \in (0, \frac{\mu}{2})$ such that $ML(a|x) = 1$ or $ML(1|x) = 1$ for almost every $x \in N(\mathcal{X}_{a,1}, 3\bar{\delta}) \cap N(\partial\Omega^*, 3\bar{\delta})$. It follows that for $(u, v, \delta) \in \partial\Omega^* \cap N(\mathcal{X}_{a,1}, \bar{\delta}) \times (-1, 1) \times (0, \bar{\delta})$, $p_\delta^{ML}(a|u + \delta v\nu_{\Omega^*}(u)) + p_\delta^{ML}(1|u + \delta v\nu_{\Omega^*}(u)) = 1$ so that $q_\delta^{ML}(u + \delta v\nu_{\Omega^*}(u)) = p_\delta^{ML}(a|u + \delta v\nu_{\Omega^*}(u))$.

For $(u, v, \delta) \in \partial\Omega^* \cap N(\mathcal{X}_{a,1}, \bar{\delta}) \times (-1, 1) \times (0, \bar{\delta})$,

$$\begin{aligned}
&p_\delta^{ML}(a|u + \delta v\nu_{\Omega^*}(u)) \\
&= \frac{\int_{B(\mathbf{0},1)} ML(a|u + \delta v\nu_{\Omega^*}(u) + \delta w)dw}{\int_{B(\mathbf{0},1)} dw} \\
&= \frac{\int_{B(\mathbf{0},1)} 1\{u + \delta v\nu_{\Omega^*}(u) + \delta w \in \Omega^*\}dw}{\mathrm{Vol}_p} \\
&= \frac{\int_{B(\mathbf{0},1)} 1\{d_{\Omega^*}^s(u + \delta(v\nu_{\Omega^*}(u) + w)) \geq 0)\}dw}{\mathrm{Vol}_p} \\
&= \frac{\int_{B(\mathbf{0},1)} 1\{d_{\Omega^*}^s(u) + \nabla d_{\Omega^*}^s(y_d(u, \delta, v, w))'\delta(v\nu_{\Omega^*}(u) + w) \geq 0\}dw}{\mathrm{Vol}_p} \\
&= \frac{\int_{B(\mathbf{0},1)} 1\{\nu_{\Omega^*}(y_d(u, \delta, v, w)) \cdot \delta(v\nu_{\Omega^*}(u) + w) \geq 0\}dw}{\mathrm{Vol}_p} \\
&= \frac{\int_{B(\mathbf{0},1)} 1\{\nu_{\Omega^*}(y_d(u, \delta, v, w)) \cdot (v\nu_{\Omega^*}(u) + w) \geq 0\}dw}{\mathrm{Vol}_p},
\end{aligned}$$

where Vol_p denotes the volume of the p-dimensional unit ball, the fourth equality follows by the mean value theorem with $y_d(u,\delta,v,w)$ on the line segment connecting u with $u+\delta(v\nu_{\Omega^*}(u)+w)$, and the second last follows since $d^s_{\Omega^*}(u)=0$ for $u\in\partial\Omega^*$ and $\nabla d^s_{\Omega^*}(x)=\nu_{\Omega^*}(x)$ for $x\in N(\partial\Omega^*,\mu)$. Since $\lim_{\delta\to 0} y_d(u,\delta,v,w)=u$ and ν_{Ω^*} is continuous,

$$\lim_{\delta\to 0}\nu_{\Omega^*}(y_d(u,\delta,v,w))\cdot(v\nu_{\Omega^*}(u)+w)=\nu_{\Omega^*}(u)\cdot(v\nu_{\Omega^*}(u)+w)=v+\nu_{\Omega^*}(u)\cdot w.$$

Therefore,

$$\lim_{\delta\to 0}1\{\nu_{\Omega^*}(y_d(u,\delta,v,w))\cdot(v\nu_{\Omega^*}(u)+w)\geq 0\}=\begin{cases}1 & \text{if } v+\nu_{\Omega^*}(u)\cdot w>0,\\ 0 & \text{if } v+\nu_{\Omega^*}(u)\cdot w<0.\end{cases}$$

By the Dominated Convergence Theorem,

$$\lim_{\delta\to 0}p^{ML}_\delta(a|u+\delta v\nu_{\Omega^*}(u))=\frac{\int_{B(\mathbf{0},1)}1\{v+\nu_{\Omega^*}(u)\cdot w>0\}dw}{\mathrm{Vol}_p}.$$

Note that the set $\{w\in B(\mathbf{0},1):v+\nu(u)\cdot w>0\}$ is a region of the p-dimensional unit ball cut off by the plane $\{w\in\mathbb{R}^p: v+\nu(u)\cdot w=0\}$. The distance from the center of the unit ball to the plane is $|v|$. Using the formula for the volume of a hyperspherical cap (see e.g. (Li 2011)), we have

$$\int_{B(\mathbf{0},1)}1\{v+\nu(u)\cdot w>0\}dw=\begin{cases}\mathrm{Vol}_p-\frac{1}{2}\mathrm{Vol}_p I_{(2(1-v)-(1-v)^2)}(\frac{p+1}{2},\frac{1}{2}) & \text{for } v\in[0,1)\\ \frac{1}{2}\mathrm{Vol}_p I_{(2(1+v)-(1+v)^2)}(\frac{p+1}{2},\frac{1}{2}) & \text{for } v\in(-1,0).\end{cases}$$

Therefore, $\lim_{\delta\to 0}p^{ML}_\delta(a|u+\delta v\nu_{\Omega^*}(u))=k(v)$. □

Step G.0.0.2. For every $(u,v,\delta)\in\partial\Omega^*\cap N(\mathcal{X}_{a,1},\bar\delta)\times(-1,1)\times(0,\bar\delta)$, $q^{ML}_\delta(u+\delta v\nu_{\Omega^*}(u))\in(0,1)$.

Proof. Fix $(u,v,\delta)\in\partial\Omega^*\cap N(\mathcal{X}_{a,1},\bar\delta)\times(-1,1)\times(0,\bar\delta)$. As discussed in G.0.0.1, $q^{ML}_\delta(u+\delta v\nu_{\Omega^*}(u))=p^{ML}_\delta(a|u+\delta v\nu_{\Omega^*}(u))$. Suppose $v=0$. By Step G.0.0.1, $p^{ML}(a|u)=\lim_{\delta'\to 0}p^{ML}_{\delta'}(a|u)=k(0)=\frac{1}{2}$. This implies that there exists $\delta'\in(0,\delta)$ such that $p^{ML}_{\delta'}(a|u)\in(0,1)$. It then follows that $0<\mathcal{L}^p(B(u,\delta')\cap\Omega^*)\leq\mathcal{L}^p(B(x,\delta)\cap\Omega^*)$ and that $0<\mathcal{L}^p(B(x,\delta')\setminus\Omega^*)\leq\mathcal{L}^p(B(x,\delta)\setminus\Omega^*)$. Therefore, $p^{ML}_\delta(a|u)=\frac{\mathcal{L}^p(B(u,\delta)\cap\Omega^*)}{\mathcal{L}^p(B(u,\delta))}\in(0,1)$.

Suppose $v\neq 0$ and let $\epsilon\in(0,\delta(1-|v|))$. Note that $B(u,\epsilon)\subset B(u+\delta v\nu_{\Omega^*}(u),\delta)$, since for any $x\in B(u,\epsilon)$, $\|u+\delta v\nu_{\Omega^*}(u)-x\|\leq\|\delta v\nu_{\Omega^*}(u)\|+\|u-x\|\leq\delta|v|+\epsilon<\delta$. Since $p^{ML}(a|u)=\frac{1}{2}$, there exists $\epsilon'\in(0,\epsilon)$ such that $p^{ML}_{\epsilon'}(a|u)\in(0,1)$. It then follows that $0<\mathcal{L}^p(B(u,\epsilon')\cap\Omega^*)\leq\mathcal{L}^p(B(u,\epsilon)\cap\Omega^*)\leq\mathcal{L}^p(B(u+\delta v\nu_{\Omega^*}(u),\delta)\cap\Omega^*)$ and that $0<\mathcal{L}^p(B(x,\epsilon')\setminus\Omega^*)\leq\mathcal{L}^p(B(x,\epsilon)\setminus\Omega^*)\leq\mathcal{L}^p(B(u+\delta v\nu_{\Omega^*}(u),\delta)\setminus\Omega^*)$. Therefore, $p^{ML}_\delta(a|u+\delta v\nu_{\Omega^*}(u))=\frac{\mathcal{L}^p(B(u+\delta v\nu_{\Omega^*}(u),\delta)\cap\Omega^*)}{\mathcal{L}^p(B(u+\delta v\nu_{\Omega^*}(u),\delta))}\in(0,1)$. □

Step G.0.0.3. Let $g:\mathbb{R}^p\to\mathbb{R}$ be a function that is bounded on $N(\partial\Omega^*,\delta')\cap N(\mathcal{X}_{a,1},\delta')$ for some $\delta'>0$. Then, for $l\geq 0$, there exist $\tilde\delta>0$ and constant $C>0$ such that

$$|\delta^{-1}E[q^{ML}_\delta(X_i)^l g(X_i)1\{q^{ML}_\delta(X_i)\in(0,1)\}1\{A_i\in\{1,a\}\}]|\leq C$$

for every $\delta\in(0,\tilde\delta)$. If g is continuous on $N(\partial\Omega^*,\delta')\cap N(\mathcal{X}_{a,1},\delta')$ for some $\delta'>0$, then

$$\begin{aligned}&\delta^{-1}E[q^{ML}_\delta(X_i)^l g(X_i)1\{q^{ML}_\delta(X_i)\in(0,1)\}1\{A_i\in\{1,a\}\}]\\&=\int_{-1}^1 k(v)^l dv\int_{\partial\Omega^*\cap\mathcal{X}_{a,1}}g(x)f_X(x)d\mathcal{H}^{p-1}(x)+o(1),\\&\delta^{-1}E[1\{A_i=a\}q^{ML}_\delta(X_i)^l g(X_i)1\{q^{ML}_\delta(X_i)\in(0,1)\}]\\&=\int_0^1 k(v)^l dv\int_{\partial\Omega^*\cap\mathcal{X}_{a,1}}g(x)f_X(x)d\mathcal{H}^{p-1}(x)+o(1)\end{aligned}$$

for $l\geq 0$.

Proof. Let $\bar\delta$ be given in Step G.0.0.1. Under Assumption 6 (i), there exists $\tilde\delta\in(0,\bar\delta)$ such that f_X is bounded and continuous on $N(\partial\Omega^*,2\tilde\delta)\cap N(\mathcal{X}_{a,1},2\tilde\delta)$. Let $\tilde\delta\in(0,\bar\delta)$ be such that both g and f_X are bounded on $N(\partial\Omega^*,2\tilde\delta)\cap N(\mathcal{X}_{a,1},2\tilde\delta)$ and such that $ML(a|x)=1$ or $ML(1|x)=1$ for almost every $x\in N(\mathcal{X}_{a,1},\tilde\delta)\cap N(\partial\Omega^*,\tilde\delta)$. Such $\tilde\delta$ exists under Assumption 6 (h) (2) and (i).

We first show that $q^{ML}_\delta(x)\in\{0,1\}$ for every $x\in\mathcal{X}_{a,1}\setminus N(\partial\Omega^*,\delta)$ for every $\delta\in(0,\tilde\delta)$. Pick $x\in\mathcal{X}_{a,1}\setminus N(\partial\Omega^*,\delta)$ and $\delta\in(0,\tilde\delta)$. Since $B(x,\delta)\cap\partial\Omega^*=\emptyset$, either $B(x,\delta)\subset\mathrm{int}(\Omega^*)$ or $B(x,\delta)\subset\mathrm{int}(\mathbb{R}^p\setminus\Omega^*)$. If $B(x,\delta)\subset\mathrm{int}(\Omega^*)$,

$q_\delta^{ML}(x) = 1$. If $B(x,\delta) \subset \mathrm{int}(\mathbb{R}^p \setminus \Omega^*)$, $q_\delta^{ML}(x) = 0$, since $ML(a|x') = 0$ for all $x' \in \mathbb{R}^p \setminus \Omega^*$ by Assumption 6 (f). Therefore, $\{x \in \mathcal{X}_{a,1} : q_\delta^{ML}(x) \in (0,1)\} \subset N(\partial\Omega^*, \delta)$ for every $\delta \in (0, \tilde{\delta})$. This implies that $ML(a|x') = 1$ or $ML(1|x') = 1$ for almost every $x' \in \{x \in \mathcal{X}_{a,1} : q_\delta^{ML}(x) \in (0,1)\}$, since $ML(a|x) = 1$ or $ML(1|x) = 1$ for almost every $x \in N(\mathcal{X}_{a,1}, \tilde{\delta}) \cap N(\partial\Omega^*, \tilde{\delta})$.

Using this result and Lemma F.5, for $\delta \in (0, \tilde{\delta})$,

$$\begin{aligned}
&\delta^{-1} E[q_\delta^{ML}(X_i)^l g(X_i) 1\{q_\delta^{ML}(X_i) \in (0,1)\} 1\{A_i \in \{1,a\}\}] \\
&= \delta^{-1} E[q_\delta^{ML}(X_i)^l g(X_i) 1\{q_\delta^{ML}(X_i) \in (0,1)\} (ML(a|X_i) + ML(1|X_i)) 1\{X_i \in \mathcal{X}_{a,1}\}] \\
&= \delta^{-1} E[q_\delta^{ML}(X_i)^l g(X_i) 1\{q_\delta^{ML}(X_i) \in (0,1)\} 1\{X_i \in \mathcal{X}_{a,1}\}] \\
&= \delta^{-1} \int q_\delta^{ML}(x)^l g(x) 1\{q_\delta^{ML}(x) \in (0,1)\} f_X(x) 1\{x \in \mathcal{X}_{a,1}\} dx \\
&= \delta^{-1} \int_{N(\partial\Omega^*, \delta)} q_\delta^{ML}(x)^l g(x) 1\{q_\delta^{ML}(x) \in (0,1)\} f_X(x) 1\{x \in \mathcal{X}_{a,1}\} dx \\
&= \delta^{-1} \int_{-\delta}^{\delta} \int_{\partial\Omega^*} q_\delta^{ML}(u + \lambda\nu_{\Omega^*}(u))^l g(u + \lambda\nu_{\Omega^*}(u)) 1\{q_\delta^{ML}(u + \lambda\nu_{\Omega^*}(u)) \in (0,1)\} \\
&\qquad \times f_X(u + \lambda\nu_{\Omega^*}(u)) 1\{u + \lambda\nu_{\Omega^*}(u) \in \mathcal{X}_{a,1}\} J_{p-1}^{\partial\Omega^*} \psi_{\Omega^*}(u,\lambda) d\mathcal{H}^{p-1}(u) d\lambda.
\end{aligned}$$

With change of variables $v = \frac{\lambda}{\delta}$, we have

$$\begin{aligned}
&\delta^{-1} E[q_\delta^{ML}(X_i)^l g(X_i) 1\{q_\delta^{ML}(X_i) \in (0,1)\} 1\{A_i \in \{1,a\}\}] \\
&= \int_{-1}^{1} \int_{\partial\Omega^*} q_\delta^{ML}(u + \delta v\nu_{\Omega^*}(u))^l 1\{q_\delta^{ML}(u + \delta v\nu_{\Omega^*}(u)) \in (0,1)\} 1\{u + \delta v\nu_{\Omega^*}(u) \in \mathcal{X}_{a,1}\} \\
&\qquad \times g(u + \delta v\nu_{\Omega^*}(u)) f_X(u + \delta v\nu_{\Omega^*}(u)) J_{p-1}^{\partial\Omega^*} \psi_{\Omega^*}(u, \delta v) d\mathcal{H}^{p-1}(u) dv.
\end{aligned}$$

For every $(u, v, \delta) \in \partial\Omega^* \setminus N(\mathcal{X}_{a,1}, \tilde{\delta}) \times (-1,1) \times (0, \tilde{\delta})$, $u + \delta v\nu_{\Omega^*}(u) \notin \mathcal{X}_{a,1}$, so

$$\begin{aligned}
&\delta^{-1} E[q_\delta^{ML}(X_i)^l g(X_i) 1\{q_\delta^{ML}(X_i) \in (0,1)\} 1\{A_i \in \{1,a\}\}] \\
&= \int_{-1}^{1} \int_{\partial\Omega^* \cap N(\mathcal{X}_{a,1}, \tilde{\delta})} q_\delta^{ML}(u + \delta v\nu_{\Omega^*}(u))^l 1\{q_\delta^{ML}(u + \delta v\nu_{\Omega^*}(u)) \in (0,1)\} \\
&\quad \times 1\{u + \delta v\nu_{\Omega^*}(u) \in \mathcal{X}_{a,1}\} g(u + \delta v\nu_{\Omega^*}(u)) f_X(u + \delta v\nu_{\Omega^*}(u)) J_{p-1}^{\partial\Omega^*} \psi_{\Omega^*}(u, \delta v) d\mathcal{H}^{p-1}(u) dv \\
&= \int_{-1}^{1} \int_{\partial\Omega^* \cap N(\mathcal{X}_{a,1}, \tilde{\delta})} q_\delta^{ML}(u + \delta v\nu_{\Omega^*}(u))^l 1\{u + \delta v\nu_{\Omega^*}(u) \in \mathcal{X}_{a,1}\} \\
&\qquad \times g(u + \delta v\nu_{\Omega^*}(u)) f_X(u + \delta v\nu_{\Omega^*}(u)) J_{p-1}^{\partial\Omega^*} \psi_{\Omega^*}(u, \delta v) d\mathcal{H}^{p-1}(u) dv
\end{aligned}$$

where the second equality follows from Step G.0.0.2. By Lemma F.5, $J_{p-1}^{\partial\Omega^*} \psi_{\Omega^*}(\cdot,\cdot)$ is bounded on $\partial\Omega^* \times (-\tilde{\delta}, \tilde{\delta})$. Since g and f_X are also bounded, for some constant $C > 0$,

$$\begin{aligned}
&|\delta^{-1} E[q_\delta^{ML}(X_i)^l g(X_i) 1\{q_\delta^{ML}(X_i) \in (0,1)\} 1\{A_i \in \{1,a\}\}]| \\
&\le C \int_{-1}^{1} \int_{\partial\Omega^* \cap N(\mathcal{X}_{a,1}, \tilde{\delta})} d\mathcal{H}^{p-1}(u) dv,
\end{aligned}$$

which is finite by Assumption 6 (h) (1).

Now suppose that g and f_X are continuous on $N(\partial\Omega^*, 2\tilde{\delta}) \cap N(\mathcal{X}_{a,1}, 2\tilde{\delta})$. We can write

$$\begin{aligned}
&\delta^{-1}E[q_\delta^{ML}(X_i)^l g(X_i)1\{q_\delta^{ML}(X_i) \in (0,1)\}1\{A_i \in \{1,a\}\}] \\
&= \int_{-1}^{1}\int_{\partial\Omega^*\cap \mathrm{int}(\mathcal{X}_{a,1})} q_\delta^{ML}(u+\delta v\nu_{\Omega^*}(u))^l 1\{u+\delta v\nu_{\Omega^*}(u) \in \mathcal{X}_{a,1}\} \\
&\qquad \times g(u+\delta v\nu_{\Omega^*}(u)) f_X(u+\delta v\nu_{\Omega^*}(u)) J_{p-1}^{\partial\Omega^*}\psi_{\Omega^*}(u,\delta v)d\mathcal{H}^{p-1}(u)dv \\
&\quad + \int_{-1}^{1}\int_{\partial\Omega^*\cap \partial\mathcal{X}_{a,1}} q_\delta^{ML}(u+\delta v\nu_{\Omega^*}(u))^l 1\{u+\delta v\nu_{\Omega^*}(u) \in \mathcal{X}_{a,1}\} \\
&\qquad \times g(u+\delta v\nu_{\Omega^*}(u)) f_X(u+\delta v\nu_{\Omega^*}(u)) J_{p-1}^{\partial\Omega^*}\psi_{\Omega^*}(u,\delta v)d\mathcal{H}^{p-1}(u)dv \\
&\quad + \int_{-1}^{1}\int_{\partial\Omega^*\cap (N(\mathcal{X}_{a,1},\tilde{\delta})\setminus \mathrm{cl}(\mathcal{X}_{a,1}))} q_\delta^{ML}(u+\delta v\nu_{\Omega^*}(u))^l 1\{u+\delta v\nu_{\Omega^*}(u) \in \mathcal{X}_{a,1}\} \\
&\qquad \times g(u+\delta v\nu_{\Omega^*}(u)) f_X(u+\delta v\nu_{\Omega^*}(u)) J_{p-1}^{\partial\Omega^*}\psi_{\Omega^*}(u,\delta v)d\mathcal{H}^{p-1}(u)dv.
\end{aligned}$$

The second term is zero by Assumption 6 (h) (1). Observe that $u+\delta v\nu_{\Omega^*}(u) \in \mathcal{X}_{a,1}$ for any sufficiently small $\delta > 0$ if $u \in \mathrm{int}(\mathcal{X}_{a,1})$ and that $u+\delta v\nu_{\Omega^*}(u) \notin \mathcal{X}_{a,1}$ for any sufficiently small $\delta > 0$ if $u \notin \mathrm{cl}(\mathcal{X}_{a,1})$. Therefore, by the Dominated Convergence Theorem,

$$\begin{aligned}
&\delta^{-1}E[q_\delta^{ML}(X_i)^l g(X_i)1\{q_\delta^{ML}(X_i) \in (0,1)\}1\{A_i \in \{1,a\}\}] \\
&\to \int_{-1}^{1}\int_{\partial\Omega^*\cap \mathrm{int}(\mathcal{X}_{a,1})} k(v)^l g(u) f_X(u) J_{p-1}^{\partial\Omega^*}\psi_{\Omega^*}(u,0)d\mathcal{H}^{p-1}(u)dv \\
&= \int_{-1}^{1} k(v)^l dv \int_{\partial\Omega^*\cap \mathcal{X}_{a,1}} g(u) f_X(u) d\mathcal{H}^{p-1}(u),
\end{aligned}$$

where we use the fact from Lemma F.5 that $J_{p-1}^{\partial\Omega^*}\psi_{\Omega^*}(u,\lambda)$ is continuous in λ and $J_{p-1}^{\partial\Omega^*}\psi_{\Omega^*}(u,0) = 1$.

Now note that $ML(a|x) = 1$ for every $x \in \Omega^*$ and $ML(a|x) = 0$ for almost every $x \in N(\mathcal{X}_{a,1}, 2\tilde{\delta}) \setminus \Omega^*$. Also, for every $(u,v,\delta) \in \partial\Omega^* \cap N(\mathcal{X}_{a,1}, \tilde{\delta}) \times (-1,1) \times (0,\tilde{\delta})$, $u+\delta v\nu_{\Omega^*}(u) \in \Omega^*$ if $v \in (0,1)$ and $u+\delta v\nu_{\Omega^*}(u) \in N(\mathcal{X}_{a,1}, 2\tilde{\delta}) \setminus \Omega^*$ if $v \in (-1,0]$. Therefore,

$$\begin{aligned}
&\delta^{-1}E[1\{A_i = a\} q_\delta^{ML}(X_i)^l g(X_i)1\{q_\delta^{ML}(X_i) \in (0,1)\}] \\
&= \delta^{-1}E[ML(a|X_i) q_\delta^{ML}(X_i)^l g(X_i)1\{q_\delta^{ML}(X_i) \in (0,1)\}1\{X_i \in \mathcal{X}_{a,1}\}] \\
&= \int_{-1}^{1}\int_{\partial\Omega^*\cap N(\mathcal{X}_{a,1},\tilde{\delta})} ML(a|u+\delta v\nu_{\Omega^*}(u)) q_\delta^{ML}(u+\delta v\nu_{\Omega^*}(u))^l 1\{u+\delta v\nu_{\Omega^*}(u) \in \mathcal{X}_{a,1}\} \\
&\qquad \times g(u+\delta v\nu_{\Omega^*}(u)) f_X(u+\delta v\nu_{\Omega^*}(u)) J_{p-1}^{\partial\Omega^*}\psi_{\Omega^*}(u,\delta v)d\mathcal{H}^{p-1}(u)dv \\
&= \int_{0}^{1}\int_{\partial\Omega^*\cap N(\mathcal{X}_{a,1},\tilde{\delta})} q_\delta^{ML}(u+\delta v\nu_{\Omega^*}(u))^l 1\{u+\delta v\nu_{\Omega^*}(u) \in \mathcal{X}_{a,1}\} \\
&\qquad \times g(u+\delta v\nu_{\Omega^*}(u)) f_X(u+\delta v\nu_{\Omega^*}(u)) J_{p-1}^{\partial\Omega^*}\psi_{\Omega^*}(u,\delta v)d\mathcal{H}^{p-1}(u)dv \\
&\to \int_{0}^{1} k(v)^l dv \int_{\partial\Omega^*\cap \mathcal{X}_{a,1}} g(u) f_X(u) d\mathcal{H}^{p-1}(u).
\end{aligned}$$

□

Step G.0.0.4. Let

$$S_{\mathbf{Z}} = \lim_{\delta\to 0}\delta^{-1}E[\mathbf{Z}_i\mathbf{Z}_i' 1\{q_\delta^{ML}(X_i) \in (0,1)\}1\{A_i \in \{1,a\}\}]$$

and

$$S_Y = \lim_{\delta\to 0}\delta^{-1}E[\mathbf{Z}_i Y_i 1\{q_\delta^{ML}(X_i) \in (0,1)\}1\{A_i \in \{1,a\}\}].$$

Then the second element of $S_{\mathbf{Z}}^{-1}S_Y$ is

$$\frac{\int_{\partial\Omega^*\cap\mathcal{X}_{a,1}} E[Y_i(a) - Y_i(1)|X_i = x] f_X(x) d\mathcal{H}^{p-1}(x)}{\int_{\partial\Omega^*\cap\mathcal{X}_{a,1}} f_X(x) d\mathcal{H}^{p-1}(x)}.$$

Under Assumption 3, this is equal to $\beta(a,1)$.

Proof. Note that

$$\begin{aligned}
&E[\mathbf{Z}_i Y_i 1\{q_\delta^{ML}(X_i) \in (0,1)\}1\{A_i \in \{1,a\}\}] \\
=& E[\mathbf{Z}_i(1\{A_i = a\}Y_i(a) + 1\{A_i = 1\}Y_i(1))1\{q_\delta^{ML}(X_i) \in (0,1)\}] \\
=& E[\mathbf{Z}_i(E[1\{A_i = a\}|X_i]E[Y_i(a)|X_i] + E[1\{A_i = 1\}|X_i]E[Y_i(1)|X_i])1\{q_\delta^{ML}(X_i) \in (0,1)\}] \\
=& E[\mathbf{Z}_i(1\{A_i = a\}E[Y_i(a)|X_i] + 1\{A_i = 1\}E[Y_i(1)|X_i])1\{q_\delta^{ML}(X_i) \in (0,1)\}],
\end{aligned}$$

where the second equality holds since A_i is independent of $Y_i(\cdot)$ conditional on X_i. By Step G.0.0.3,

$$S_{\mathbf{Z}} = \bar{f}_X \begin{bmatrix} 2 & 1 & \int_{-1}^1 k(v)dv \\ 1 & 1 & \int_0^1 k(v)dv \\ \int_{-1}^1 k(v)dv & \int_0^1 k(v)dv & \int_{-1}^1 k(v)^2 dv \end{bmatrix},$$

where $\bar{f}_X = \int_{\partial\Omega^* \cap \mathcal{X}_{a,1}} f_X(x) d\mathcal{H}^{p-1}(x)$, and

$$S_Y = \begin{bmatrix} \int_{\partial\Omega^* \cap \mathcal{X}_{a,1}} E[Y_i(a) + Y_i(1)|X_i = x] f_X(x) d\mathcal{H}^{p-1}(x) \\ \int_{\partial\Omega^* \cap \mathcal{X}_{a,1}} E[Y_i(a)|X_i = x] f_X(x) d\mathcal{H}^{p-1}(x) \\ \int_{\partial\Omega^* \cap \mathcal{X}_{a,1}} (\int_0^1 k(v)dv E[Y_i(a)|X_i = x] + \int_{-1}^0 k(v)dv E[Y_i(1)|X_i = x]) f_X(x) d\mathcal{H}^{p-1}(x) \end{bmatrix}.$$

After a few lines of algebra, we have

$$\det(S_{\mathbf{Z}}) = \bar{f}_X^{-1}\left(\int_{-1}^0 (k(v) - \int_{-1}^0 k(s)ds)^2 dv + \int_0^1 (k(v) - \int_0^1 k(s)ds)^2 dv\right),$$

which is nonzero under Assumption 6 (h) (1). After another few lines of algebra, we obtain that the second element of $S_{\mathbf{Z}}^{-1} S_Y$ is

$$\frac{\int_{\partial\Omega^* \cap \mathcal{X}_{a,1}} E[Y_i(a) - Y_i(1)|X_i = x] f_X(x) d\mathcal{H}^{p-1}(x)}{\int_{\partial\Omega^* \cap \mathcal{X}_{a,1}} f_X(x) d\mathcal{H}^{p-1}(x)} = \beta(a,1).$$

Note that if Assumption 3 does not hold, the left-hand side still represents the mean reward difference for the subpopulation on the boundary $\partial\Omega^* \cap \mathcal{X}_{a,1}$. □

Step G.0.0.5. If $n\delta_n \to \infty$ as $n \to \infty$, then $\hat{\beta}_a \xrightarrow{p} \beta(a,1)$.

Proof. It suffices to verify that the variance of each element of $\frac{1}{n\delta_n}\sum_{i=1}^n \mathbf{Z}_i\mathbf{Z}_i' I_i 1\{A_i \in \{1,a\}\}$ and $\frac{1}{n\delta_n}\sum_{i=1}^n \mathbf{Z}_i Y I_i 1\{A_i \in \{1,a\}\}$ is $o(1)$. Here, we only verify that $\mathrm{Var}(\frac{1}{n\delta_n}\sum_{i=1}^n q_{\delta_n}^{ML}(X_i) Y_i I_i 1\{A_i \in \{1,a\}\}) = o(1)$. Note that

$$\begin{aligned}
&E[Y_i^2 1\{A_i \in \{1,a\}\}|X_i] = E[1\{A_i = a\}Y_i(a)^2 + 1\{A_i = 1\}Y_i(1)^2|X_i] \\
&\le E[Y_i(a)^2 + Y_i(1)^2|X_i]E[1\{A_i \in \{1,a\}\}|X_i].
\end{aligned}$$

Under Assumption 6 (i), there exists $\delta' > 0$ such that $E[Y_i(a)^2 + Y_i(1)^2|X_i]$ is bounded on $N(\partial\Omega^*, \delta')$. We have

$$\begin{aligned}
&\mathrm{Var}\left(\frac{1}{n\delta_n}\sum_{i=1}^n q_{\delta_n}^{ML}(X_i) Y_i I_i 1\{A_i \in \{1,a\}\}\right) \\
\le\ & \frac{1}{n\delta_n}\delta_n^{-1} E[q_{\delta_n}^{ML}(X_i)^2 Y_i^2 I_i 1\{A_i \in \{1,a\}\}] \\
=\ & \frac{1}{n\delta_n}\delta_n^{-1} E[q_{\delta_n}^{ML}(X_i)^2 E[Y_i(a)^2 + Y_i(1)^2|X_i] I_i 1\{A_i \in \{1,a\}\}] \\
\le\ & \frac{1}{n\delta_n} C
\end{aligned}$$

for some $C > 0$, where the last inequality follows from Step G.0.0.3. The conclusion follows since $n\delta_n \to \infty$. □