Offline Dynamic Pricing under Covariate Shift and Local Differential Privacy via Twofold Pessimism

Jongmin Mun, Xiaocong Xu, Yingying Fan
University of Southern California
{jongmin.mun, xuxiaoco, fanyingy}@marshall.usc.edu

Abstract

We study pricing policy learning from batched contextual bandit data under market shift and privacy protection. Market shift is modeled as covariate shift, where the relationship among treatments, features, and rewards remains invariant, while privacy is enforced through local differential privacy (LDP), which perturbs each data point before use. Viewing the off-policy setting, covariate shift, and LDP collectively as forms of distributional shift, we develop a policy learning algorithm based on a unified pessimism principle that addresses all three shifts. Without privacy, we estimate the conditional reward via nonparametric regression and quantify its variance to construct a pessimistic estimator, yielding a policy with minimax-optimal decision error. Under LDP, we apply the Laplace mechanism and adjust the pessimistic estimator to account for additional uncertainty from privacy noise. The resulting doubly pessimistic objective is then optimized to determine the final pricing policy.

1 Introduction

Data-generating Process We study an observational dataset collected from interactions with a dynamic pricing system, following standard formulations in offline policy learning [3, 8, 10, 15, 17]. For each interaction, the system observes an exogenous context vector $\mathbf{X}_i^P \in \mathcal{X} = [0,1]^d$ encoding the product's characteristics, the customers' information and some other confounding factors, for example, economic indicator, weather, competitors' treatments [21, 5, 24]. The system assigns a treatment (action) $T_i^P \in \mathcal{T} = [0,1]$ based on an unknown logging policy, characterized by the generalized propensity score $\pi_{\text{off}}(\cdot \mid \mathbf{X}_i^P)$, which generalizes the discrete propensity $\Pr(T_i^P = a)$ to continuous treatments [7, 9]. In loan pricing, T_i^P may represent an interest rate [12], whereas in settings with a fixed product type, it can denote a $(1 - T_i^P)\%$ discount from a baseline price [6]. Let $Y(t) \in \mathcal{Y} := [0,1]$ denote the potential outcome (reward) under treatment t, following Rubin's potential outcome framework [22]. We assume that the features, treatments, and potential rewards, denoted $(\mathbf{X}, T, \{Y(t)\}_{t\in\mathcal{T}})$, follow an unknown joint distribution D^P. The observed reward is $Y_i^P := Y(T_i^P)$. Thus, the observational dataset $\mathcal{D}_n = \{(\mathbf{X}_i^P, T_i^P, Y_i^P)\}_{i=1}^n$ consists of i.i.d. samples from D^P.

We define the conditional average outcome, or individualized dose–response function (IDRF), as $f(\mathbf{x}, t) := \mathbb{E}\big[Y(t) \mid \mathbf{X} = \mathbf{x}\big]$. We impose the following structural assumptions on D^P. First, we assume that the potential reward is sub-Gaussian and IDRF is continuous.

Assumption 1 (Sub-Gaussianity and β-Hölder continuity). *The potential reward $Y(t)$ is 1-sub-Gaussian. Moreover, corresponding IDRF is β-Hölder continuous for some $\beta \in (0,1]$. That is, there exists a constant $C_{\mathrm{H}} > 0$ such that for all $(\mathbf{x}_1, t_1), (\mathbf{x}_2, t_2) \in \mathcal{X} \times \mathcal{T}$,*

$$|f(\mathbf{x}_1, t_1) - f(\mathbf{x}_2, t_2)| \;\leq\; C_{\mathrm{H}} \, \|(\mathbf{x}_1, t_1) - (\mathbf{x}_2, t_2)\|_\infty^\beta.$$

39th Conference on Neural Information Processing Systems (NeurIPS 2025) Workshop: MLxOR: Mathematical Foundations and Operational Integration of Machine Learning for Uncertainty-Aware Decision-Making.

This assumption on IDRF is weaker than those in earlier works on private dynamic pricing [25, 16, 6] and causal dynamic pricing [24]. Moreover, continuity of the IDRF is considered a desirable property when estimating the causal effects of continuous treatments [18]. To ensure that IDRF coincides with $\mathbb{E}[Y_i^P | X_i^P = \mathbf{x}, T_i^P = t]$, we further assume weak unconfoundedness of the logging policy [7].

Assumption 2 (Weak unconfoundedness). $Y(t) \perp\!\!\!\perp T_i^P \;\big|\; \mathbf{X}_i^P$ *for all* $t \in \mathcal{T}$*.*

Off-policy Learning under Covariate Shift and Local Differential Privacy We aim to learn a policy $\hat{\pi} : \mathcal{X} \to \mathcal{T}$ from observational data under two types of distributional shift: *covariate shift* and *local differential privacy*. Let D^Q denote the distribution of $\big(\mathbf{X}^Q, Y(\hat{\pi}(\mathbf{X}^Q))\big)$. Covariate shift arises when the marginal distribution of features under D^Q, denoted Q, differs from that under the historical data D^P, denoted P, while the IDRF remains unchanged [23, 14]. This typically occurs when a product enters a new market. To protect the privacy of each interaction, each observation is randomly perturbed through a privacy mechanism represented by a Markov kernel $M_i(\cdot \mid \mathbf{x}, t, y)$, where the magnitude of the perturbation is controlled by the privacy parameter ε. This mechanism ensures *local differential privacy*, formally defined in Definition 1 (Section 3). The resulting observational data are given by $\widetilde{\mathcal{D}}_n := \{V_i\}_{i=1}^n = \big\{M_i(\mathbf{X}_i^P, T_i^P, Y_i^P)\big\}_{i=1}^n$. After the learned policy is deployed, the system directly observes the target features $\mathbf{X}^Q$ without privacy protection, following Chen et al. [4, 6]. The goal of policy learning is to maximize the *policy value*, defined as the expected reward:

$$V(\pi) := \mathbb{E}\big[Y\big(\pi(\mathbf{X}^Q)\big)\big] = \mathbb{E}\big[\mathbb{E}\big[Y\big(\pi(\mathbf{X}^Q)\big) \,|\, \mathbf{X}^Q\big]\big] = \mathbb{E}\left[f\big(\mathbf{X}^Q, \pi(\mathbf{X}^Q)\big)\right], \tag{1}$$

where the expectation is taken with respect to the randomness in D^Q, D^P, π_{off}, and $M_1, \ldots, M_n$.

Previous Approaches Approaches to off-policy learning under unconfoundedness can be broadly categorized into regression-based methods [20], reweighting-based methods [13], and doubly robust methods that combine the two [1, 11]. The regression-based approach estimates the IDRF through regression, but the resulting policy learning algorithm may not be consistent if the IDRF is misspecified. The reweighting and doubly robust approaches help mitigate this issue, and these methods are widely adopted in existing literature on off-policy learning from observational data under distribution shift [26, 29].

Our Contribution This paper adopts a regression-based approach but, instead of mitigating IDRF misspecification, we leverage it to guide policy learning. Assuming Hölder continuity of the IDRF, we estimate it using nonparametric regression, which provides effective bias control at the cost of higher variance. This variance varies across treatments, being larger for those underexplored by the logging policy or source covariate distribution P. We formalize this with a pointwise error bound for the IDRF estimator and use it to construct a pessimistic estimate. For a new target context $\mathbf{X}^Q$, our policy selects the treatment that maximizes this pessimistic predicted IDRF, naturally discouraging underexplored treatments with wider confidence intervals. We further extend this framework to the local differential privacy setting.

2 Pricing policy without privacy

We present our policy in the non-private setting ($\varepsilon = \infty$), where $M_1, \ldots, M_n$ are identity mappings and thus $\widetilde{\mathcal{D}}_n = \mathcal{D}_n$. We first introduce two notations for radius-h neighborhoods corresponding to a target context $\mathbf{x}^Q \in \mathcal{X}$ and a candidate treatment $t \in \mathcal{T}$:

$$I_{h,i}(\mathbf{x}^Q, t) := \mathbb{1}\{(\|\mathbf{X}_i^P - \mathbf{x}^Q\|_\infty \vee |t_i^P - t|) \leq h\}, \quad N_h(\mathbf{x}^Q, t) := \sum_{i=1}^n I_{h,i}(\mathbf{x}^Q, t),\; i = 1, \ldots, n.$$

We then define the pessimistic estimator for f as follows:

Proposition 1 (Pessimistic IDRF estimator). *Let* $h > 0$ *be a bandwidth parameter,* $\beta > 0$ *the Hölder smoothness constant (with corresponding constant* C_{H}*),* $\mathbf{x}^Q \in \mathcal{X}$ *a target context, and* $t \in \mathcal{T}$ *a candidate treatment. Define the pessimistic estimator of the reward function as*

$$\tilde{f}_h(\mathbf{x}^Q, t) := \left(\frac{\sum_{i=1}^n Y_i I_{h,i}(\mathbf{x}^Q, t)}{N_h(\mathbf{x}^Q, t)} - \sqrt{\frac{2 \log 2n}{N_h(\mathbf{x}^Q, t)}} - C_{\mathrm{H}} h^\beta \right) \mathbb{1}\big(N_h(\mathbf{x}^Q, t) > 0\big). \tag{2}$$

Under Assumptions 1 and 2, with probability at least $1 - 1/n$*, we have* $\tilde{f}_h(\mathbf{x}^Q, t) \leq f(\mathbf{x}^Q, t)$*.*

The proof of Proposition 1 is given in Appendix C. Building on this pessimistic estimator, our policy selects the treatment $\widehat{\pi}_h(\mathbf{x}^{\mathbf{Q}})$ as the smallest maximizer:

$$\widehat{\pi}_h(\mathbf{x}^Q) := \min\left\{t' \in \arg\max_{t\in\mathcal{T}} \tilde{f}_h(\mathbf{x}^Q, t)\right\}. \tag{3}$$

Since $\tilde{f}_h(\mathbf{x}^Q, t)$ is piecewise constant, changing only at the n indicator thresholds or at the domain endpoints, it has at most $2n + 2$ segments for a fixed $\mathbf{x}^Q$. Hence, computing $\widehat{\pi}_h(\mathbf{x}^Q)$ reduces to comparing its values across these segments. The full procedure is summarized in Algorithm 2.

We now provide a theoretical guarantee for the proposed policy learning algorithm. We begin by introducing the assumptions required to establish this guarantee. Intuitively, successful transfer relies on the similarity between the source data generator and the target data generator, as well as the effectiveness of the offline policy π_{off} in exploring optimal treatments. Accordingly, the minimax rate depends on two key factors: the overlap between the source and target context distributions, P and Q, and the extent of exploration performed by the offline policy. We formalize the first quantity, which was initially introduced by Kpotufe and Martinet [14].

Assumption 3 (Transfer exponent)**.** *There exist constants $\alpha \geq 0$, $\bar{h} > 0$, and $C_\alpha > 0$ such that for all $x \in \operatorname{supp}(Q)$ and all $h \in (0, \bar{h}]$,*

$$P\big(B_h(\mathbf{x})\big) \geq C_\alpha h^\alpha Q\big(B_h(\mathbf{x})\big). \tag{4}$$

We denote this condition as $\kappa_{\bar{h}}(P, Q) \leq \alpha$. For brevity, when the context is clear, we simplify the notation to $\kappa(P, Q) \equiv \kappa_{\bar{h}}(P, Q)$.

Note that any pair (P, Q) satisfies the above with $\alpha = \infty$. For treatments, however, we do not impose a strict transfer exponent. Instead, we assume only that the offline policy π_{off} under P frequently selects near-optimal treatments. This is a weaker assumption than in prior work in online setting [28], which requires exploration of the entire treatment range.

Assumption 4 (Near-optimal treatment exploration)**.** *Let μ denote the joint distribution of source context-treatment pairs $(\mathbf{X}^P, t^P)$. There exists a constant $\zeta \in [0, 1]$ such that for each $\mathbf{x} \in \operatorname{supp}(P)$, there is at least one optimal treatment, denoted $t^\dagger(\mathbf{x}) \in \arg\max_{t\in\mathcal{T}} f(\mathbf{x}, t)$ for which:*

$$\inf_{h\in(0,1/2]} \frac{\mu\big(B_h(\mathbf{x}) \times \big[t^\dagger(\mathbf{x}) - h,\ t^\dagger(\mathbf{x}) + h\big]\big)}{2hP\big(B_h(\mathbf{x})\big)} \ \geq\ \zeta. \tag{5}$$

We now show that the pricing policy (3) achieves a minimax-optimal suboptimality gap.

Theorem 2. *Let $\pi^* \in \arg\max_\pi V(\pi)$. Under Assumptions 1 - 4, there exist $\bar{h}$ and h such that,*

$$V(\pi^*) - V(\hat{\pi}_h) = \tilde{O}\big((\zeta n)^{-\frac{\beta}{2\beta+\alpha+d+1}}\big), \tag{6}$$

for sufficiently large n, where $\tilde{O}(\cdot)$ hides logarithmic factors. Moreover, if $\zeta n > 1$, then for any policy π, there exists an instance satisfying Assumptions 1–4 such that

$$V(\pi^*) - V(\pi) = \Omega\big((\zeta n)^{-\frac{\beta}{2\beta+\alpha+d+1}}\big). \tag{7}$$

The proof of Theorem 2 is provided in Appendix D.

3 Pricing policy with privacy

We now extend the pricing policy (3) to the local differential privacy setting. We begin by formally defining local differential privacy.

Definition 1 (Local differential privacy)**.** *Given a privacy level $\varepsilon > 0$, let (X_i^P, t_i^P, Y_i^P) and V_i be random elements mapped to measurable spaces $(\mathcal{X} \times \mathcal{T} \times \mathcal{Y}, \mathcal{F})$ and $(\mathcal{V}_i, \mathcal{B}_i)$, respectively, for each $i = 1, \ldots, n$. Then V_i is an ε-local differentially private (ε-LDP) view of (X_i^P, t_i^P, Y_i^P) if there exists a privacy mechanism $M_i(\cdot \mid \cdot)$, which is a bivariate function on $\mathcal{B}_i \times (\mathcal{X} \times \mathcal{T} \times \mathcal{Y})$ such that:*

1. *For any $(x, p, y) \in \mathcal{X} \times \mathcal{T} \times \mathcal{Y}$, $M_i(\cdot \mid x, p, y)$ is a conditional distribution of V_i given $(X_i^P, t_i^P, Y_i^P) = (x, p, y)$,*

2. *For any $A \in \mathcal{B}_i$, $(x, p, y) \mapsto M_i(A \,|\, x, p, y)$ is a measurable function on $\mathcal{X} \times \mathcal{T} \times \mathcal{Y}$, and*
3. *For any (x, p, y) and (x', t', y') in $\mathcal{X} \times \mathcal{T} \times \mathcal{Y}$ and $A \in \mathcal{B}_i$, the inequality $M_i(A|x, p, y) \leq e^{\varepsilon} M_i(A|x', t', y')$ holds.*

Privacy mechanism. Given a bandwidth h, let $K := \lfloor 1/h^{d+1} \rfloor$. Partition $\mathcal{X} \times \mathcal{T}$ into hypercubes $A_{h,1}, \ldots, A_{h,K}$ of side length h. Each seller $i \leq n$ privatizes their datapoint $(\mathbf{X}_i^P, t_i^P)$ using the following protocol: (i) each context-treatment pair $(\mathbf{X}_i^P, t_i^P)$ is encoded as a K-dimensional one-hot vector W_i, where the jth entry equals 1 if and only if $(\mathbf{X}_i^P, t_i^P) \in A_{h,j}$. (ii) independent Laplace noise is added to each entry of W_i and $Y_i^P W_i$. The variance of the Laplace noise are both determined by the privacy parameter ε. See Algorithm 1 for the complete procedure. The proof of ε -LDP guarantee of Algorithm 1 is provided in Proposition 1 of Berrett et al. [2].

Algorithm 1 Privacy mechanism for ith observation

```
Require: $i$'s raw obsercation $(X_i^P, t_i^P, Y_i^P)$, bandwidth $h > 0$, privacy budget $\varepsilon > 0$
$K \leftarrow \lfloor 1/h^{(d+1)} \rfloor$                                          ▷ Number of bins
$j \leftarrow$ index such that $(X_i^P, t_i^P) \in A_{h,j}$
$W_i \leftarrow K$-dimensional one-hot vector of length $K$ with 1 at position $j$     ▷ Binning
for $j = 1, \ldots, K$ do                                                 ▷ Laplace mechanism
    draw $\xi_{i,j}, \zeta_{i,j} \overset{\text{iid}}{\sim}$ standard Laplace
    $Z_{i,j} \leftarrow Y_i W_{i,j} + (4\sqrt{2}/\varepsilon)\zeta_{i,j}$
    $W_{i,j} \leftarrow W_{i,j} + (4\sqrt{2}/\varepsilon)\xi_{i,j}$
return $(W_i, Z_i) \in \mathbb{R}^K \times \mathbb{R}^K$
```

Twofold pessimistic IDRF estimator. The server receives $\widetilde{\mathcal{D}}_n := (W_i, Z_i)_{i=1}^n$ privatized by Algorithm 1 and constructs a high-probability lower bound of the true reward as follows:

Proposition 3. *Let $\varepsilon > 0$ be a privacy parameter, $h > 0$ a bandwidth parameter, $\beta > 0$ the Hölder smoothness constant, $\mathbf{x}^Q \in \mathcal{X}$ a target context, and $t \in \mathcal{T}$ a candidate treatment. If $(\mathbf{x}^Q, t) \in A_{h,j}$, given an ε-LDP dataset generated by Algorithm 1, define the noisy empirical measures and a cutoff*

$$\tilde{\mu}_n(A_{h,j}) := \frac{1}{n}\sum_{i=1}^{n} W_{i,j}, \quad \tilde{\nu}_n(A_{h,j}) := \frac{1}{n}\sum_{i=1}^{n} Z_{i,j}, \quad t_{\varepsilon} := \frac{16}{\varepsilon}\sqrt{\frac{\log n}{n}}.$$

Then, define a private pessimistic estimator of the pessimistic reward estimator in Proposition 1 as

$$\tilde{f}_{h,\mathrm{DP}}(\mathbf{x}^Q, t) := \left(\frac{\tilde{\nu}_n(A_{h,j}) - t_{\varepsilon}}{\tilde{\mu}_n(A_{h,j}) + t_{\varepsilon}} - \sqrt{\frac{(2\log 2n)/n}{\tilde{\mu}_n(A_{h,j}) - t_{\varepsilon}}} - Lh^{\beta}\right)\mathbb{1}\left(\tilde{\mu}_n(A_{h,j}) \geq t_{\varepsilon}\right).$$

Under Assumptions 1 and 2, with probability at least $1 - 4/n$, we have $\tilde{f}_{h,\mathrm{DP}}(\mathbf{x}^Q, t) \leq f(\mathbf{x}^Q, t)$.

Proof of Proposition 3 is provided in Appendix E.1.

Pricing policy. To determine the treatment, we first define the treatment gridpoints, with each point representing the treatment of its bin, as

$$\mathcal{S}_{\mathcal{T},h} = \left\{(k - \tfrac{1}{2})h : \, k \in \left[\left\lfloor \tfrac{1}{h} \right\rfloor\right]\right\} = \left\{\ddot{t}_1, \ldots, \ddot{t}_{\lfloor 1/h \rfloor}\right\}. \tag{8}$$

For a given context $\mathbf{x}^Q$, our policy sets the treatment to the smallest gridpoint that maximizes the pessimistic reward estimate $\tilde{f}_h(\mathbf{x}^Q, t)$:

$$\widehat{\pi}_{h,\mathrm{DP}}(\mathbf{x}^Q) := \min\left\{t' \in \arg\max_{\ddot{t} \in \mathcal{S}_{\mathcal{T},h}} \tilde{f}_{h,\mathrm{DP}}(\mathbf{x}^Q, \ddot{t})\right\}. \tag{9}$$

Upon receiving the privatized dataset, $\tilde{f}_{h,\mathrm{DP}}(\mathbf{x}^Q, t)$ is a piecewise-constant function with $\lfloor \frac{1}{h} \rfloor$ segments for each fixed $\mathbf{x}^Q$. Thus, determining $\widehat{\pi}_{h,\mathrm{DP}}(\mathbf{x}^Q)$ only requires comparing the function values across these segments. The complete procedure is summarized in Algorithm 3 in Appendix 3.

References

[1] Athey, S. and Wager, S. (2021). Policy learning with observational data. *Econometrica*, 89(1):133–161.

[2] Berrett, T. B., Györfi, L., and Walk, H. (2021). Strongly universally consistent nonparametric regression and classification with privatised data. *Electronic Journal of Statistics*, 15(1):2430–2453.

[3] Cai, C., Cai, T. T., and Li, H. (2024). Transfer learning for contextual multi-armed bandits. *The Annals of Statistics*, 52(1):207–232.

[4] Chen, X., Miao, S., and Wang, Y. (2023). Differential privacy in personalized pricing with nonparametric demand models. *Operations Research*, 71(2):581–602.

[5] Chen, X., Owen, Z., Pixton, C., and Simchi-Levi, D. (2022a). A statistical learning approach to personalization in revenue management. *Management Science*, 68(3):1923–1937.

[6] Chen, X., Simchi-Levi, D., and Wang, Y. (2022b). Privacy-preserving dynamic personalized pricing with demand learning. *Management Science*, 68(7):4878–4898.

[7] Hirano, K. and Imbens, G. W. (2004). The Propensity Score with Continuous Treatments. In *Applied Bayesian Modeling and Causal Inference from Incomplete-Data Perspectives*, chapter 7, pages 73–84. John Wiley & Sons, Ltd.

[8] Huang, R., Zhang, H., Melis, L., Shen, M., Hejazinia, M., and Yang, J. (2023). Federated linear contextual bandits with user-level differential privacy. In *International Conference on Machine Learning (ICML)*, pages 14060–14095.

[9] Imai, K. and van Dyk, D. A. (2004). Causal inference with general treatment regimes: Generalizing the propensity score. *Journal of the American Statistical Association*, 99(467):854–866.

[10] Kallus, N. and Zhou, A. (2018a). Policy Evaluation and Optimization with Continuous Treatments. In *International Conference on Artificial Intelligence and Statistics (AISTATS)*, pages 1243–1251. PMLR.

[11] Kallus, N. and Zhou, A. (2018b). Policy evaluation and optimization with continuous treatments. *International Conference on Artificial Intelligence and Statistics (AISTATS)*, pages 1243–1251.

[12] Keskin, N. B. and Zeevi, A. (2014). Dynamic Pricing with an Unknown Demand Model: Asymptotically Optimal Semi-Myopic Policies. *Operations Research*, 62(5):1142–1167.

[13] Kitagawa, T. and Tetenov, A. (2018). Who should be treated? Empirical welfare maximization methods for treatment choice. *Econometrica*, 86(2):591–616.

[14] Kpotufe, S. and Martinet, G. (2021). Marginal singularity and the benefits of labels in covariate-shift. *The Annals of Statistics*, 49(6):3299–3323.

[15] Lange, S., Gabel, T., and Riedmiller, M. (2012). Batch reinforcement learning. In *Reinforcement Learning*, pages 45–73. Springer, Berlin, Heidelberg.

[16] Lei, Y. M., Miao, S., and Momot, R. (2024). Privacy-preserving personalized revenue management. *Management Science*, 70(7):4875–4892.

[17] Levine, S., Kumar, A., Tucker, G., and Fu, J. (2020). Offline reinforcement learning: Tutorial, review, and perspectives on open problems. *arXiv preprint arXiv:2005.01643*.

[18] Nie, L., Ye, M., Liu, Q., and Nicolae, D. (2020). VCNet and Functional Targeted Regularization For Learning Causal Effects of Continuous Treatments. *International Conference on Learning Representations (ICLR)*.

[19] Pathak, R., Ma, C., and Wainwright, M. (2022). A new similarity measure for covariate shift with applications to nonparametric regression. In *International Conference on Machine Learning (ICML)*, pages 17517–17530.

[20] Qian, M. and Murphy, S. A. (2011). Performance guarantees for individualized treatment rules. *The Annals of Statistics*, 39(2):1180–1210.

[21] Qiang, S. and Bayati, M. (2016). Dynamic pricing with demand covariates. *arXiv preprint arXiv:1604.07463*.

[22] Rubin, D. B. (1984). Bayesianly justifiable and relevant frequency calculations for the applied statistician. *The Annals of Statistics*, 12(4):1151–1172.

[23] Shimodaira, H. (2000). Improving predictive inference under covariate shift by weighting the log-likelihood function. *Journal of Statistical Planning and Inference*, 90(2):227–244.

[24] Simchi-Levi, D. and Wang, C. (2025). Pricing experimental design: Causal effect, expected revenue and tail risk. *Management Science*.

[25] Song, B. and Jian, S. (2025). Balancing privacy and revenue: A differentially private dynamic pricing algorithm for ride-hailing. *Transportation Research Part E: Logistics and Transportation Review*, 203:104310.

[26] Uehara, M., Kato, M., and Yasui, S. (2020). Off-policy evaluation and learning for external validity under a covariate shift. *Advances in Neural Information Processing Systems*, 33:49–61.

[27] Wainwright, M. J. (2019). *High-Dimensional Statistics: A Non-Asymptotic Viewpoint*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, Cambridge.

[28] Wang, F., Jiang, F., Zhao, Z., and Yu, Y. (2025). Transfer learning for nonparametric contextual dynamic pricing. In *International Conference on Machine Learning (ICML)*.

[29] Zhao, P., Josse, J., and Yang, S. (2025). Efficient and robust transfer learning of optimal individualized treatment regimes with right-censored survival data. *Journal of Machine Learning Research*, 26(48):1–54.

A Algorithms

Algorithm 2 Non-private Off-policy Learning Algorithm

```
Require: Observational dataset $\mathcal{D}_n = \{(\mathbf{X}_i^P, t_i^P, Y_i^P)\}_{i=1}^n$, target context $\mathbf{x}^Q \in [0,1]^d$, bandwidth $h > 0$, Hölder exponent $\beta \in (0,1]$
$\mathcal{G} \leftarrow \{0,1\}$                                        ▷ treatment grid
for i=1, ..., n do
    $\mathcal{G} \leftarrow \mathcal{G} \cup \{\max(0, t_i^P - h), \min(1, t_i^P + h)\}$
LCB_max $\leftarrow -\infty$
$\widehat{\pi}_h(\mathbf{x}^Q) \leftarrow 0$
for $\ddot{t} \in \mathcal{G}$ do
    LCB_now $\leftarrow \tilde{f}(\mathbf{x}^Q, \ddot{t})$                      ▷ (2)
    if LCB_now $>$ LCB_max then
        LCB_max $\leftarrow$ LCB_now
        $\widehat{\pi}_h(\mathbf{x}^Q) \leftarrow \ddot{t}$
    else if LCB_now $=$ LCB_max and $\ddot{t} < \widehat{\pi}_h(\mathbf{x}^Q)$ then
        $\widehat{\pi}_h(\mathbf{x}) \leftarrow \ddot{t}$
return $\widehat{\pi}_h(\mathbf{x})$
```

Algorithm 3 Private pricing policy

```
Require: Target context $\mathbf{x}^Q \in \mathcal{X} = [0,1]^d$, Privatized observational dataset $\{(W_i, Z_i)\}_{i=1}^n$ processed by Algorithm 1, Hölder exponent $\beta \in (0,1]$
max_reward_now $\leftarrow 0$
$\widehat{\pi}_{h,\mathrm{DP}}(\mathbf{x}^Q) \leftarrow 1$
for $j = 1, \ldots, \lfloor 1/h \rfloor$ do                ▷ Minimum argmax procedure in (9):
    $\ddot{t}_j \leftarrow (j - 0.5)h$                             ▷ treatment candidate (8)
    if $\tilde{f}_{h,\mathrm{DP}}(\mathbf{x}^Q, \ddot{t}_j) >$ max_now then
        max_reward_now $\leftarrow \tilde{f}_{h,\mathrm{DP}}(\mathbf{x}^Q, \ddot{t}_j)$           ▷ Proposition 3
        $\widehat{\pi}_{h,\mathrm{DP}}(\mathbf{x}^Q) \leftarrow \ddot{t}_j$
Return $\widehat{\pi}_{h,\mathrm{DP}}(\mathbf{x}^Q)$
```

B Technical lemmas and definitions

The following lemma provides a bound on the expectation of the reciprocal of a binomial random variable, restricted to the event that it is positive.

Lemma 4 (Lemma 6 in 19). *Let n be a positive integer and $p \in (0,1)$. Suppose $U \sim \mathrm{Bin}(n,p)$. Then,*

$$\mathbb{E}\left[\frac{1}{U} \cdot \mathbb{1}\{U > 0\}\right] \leq \frac{4}{\mathbb{E}[U]}.$$

Lemma 5 (Lemma 15 in Cai et al. 3). *For any $a, b \in [0,1]$, let Bern(a) and Bern(b) denote two Bernoulli distributions with parameters a and b, respectively. Then one has*

$$\mathit{KL}(\mathit{Bern}(a) \| \mathit{Bern}(b)) \leq \frac{(a-b)^2}{b(1-b)}. \tag{10}$$

In addition, if $|b - 1/2| \leq 1/4$, then one further has

$$\mathit{KL}(\mathit{Bern}(a) \| \mathit{Bern}(b)) \leq 8(a-b)^2. \tag{11}$$

C Proof of Proposition 1

In the definition of the pessimistic estimator $\tilde{f}_h(\mathbf{x}^Q, t)$, recalled below:

$$\tilde{f}_h(\mathbf{x}^Q, t) := \left(\frac{\sum_{i=1}^n Y_i I_{h,i}(\mathbf{x}^Q, t)}{N_h(\mathbf{x}^Q, t)} - \sqrt{\frac{2 \log 2n}{N_h(\mathbf{x}^Q, t)}} - C_{\mathrm{H}} h^\beta \right) \mathbb{1}\big(N_h(\mathbf{x}^Q, t) > 0\big), \tag{12}$$

let us denote the first term as

$$\widehat{f}_h(\mathbf{x}, p) := \frac{\sum_{i=1}^n Y_i I_{h,i}(\mathbf{x}^Q, t)}{N_h(\mathbf{x}^Q, t)} \mathbb{1}\big(N_h(\mathbf{x}^Q, t) > 0\big) \tag{13}$$

This quantity can be interpreted as an empirical reward function based on the Nadaraya–Watson (NW) estimator. By construction, $\widehat{f}_h(\mathbf{x}, p) = 0$ whenever no source data points lie in the specified neighborhood. In the next lemma, we provide a high-probability pointwise error bound for $\widehat{f}h(\mathbf{x}, p)$.

Proposition 6 (Pointwise error bound)**.** *Suppose Assumptions 1 and 2 hold. Fix the bandwidth $h > 0$. Then, for any given target context $\mathbf{x}^Q \in \mathcal{X}$ and treatment candidate $t \in \mathcal{T}$, with probability at least $1 - 1/n$, the following bound holds:*

$$\big|\widehat{f}_h(\mathbf{x}^Q, t) - f(\mathbf{x}^Q, t)\big| \le \mathbb{1}\{N_h(\mathbf{x}^Q, t) = 0\} + \mathbb{1}\{N_h(\mathbf{x}^Q, t) > 0\} \left(L h^\beta + \sqrt{\frac{2 \log 2n}{N_h(\mathbf{x}^Q, t)}} \right).$$

The proof of Proposition 6 is provided in Appendix C.1.

Proof. We distinguish between the two cases $N_h(\mathbf{x}, p) = 0$ and $N_h(\mathbf{x}, p) > 0$.

Case 1: $N_h(\mathbf{x}, p) = 0$. In this case, the pessimistic estimator evaluates to $\widehat{f}_h(\mathbf{x}^Q, t) = 0$. Since the reward function takes values in $\mathcal{Y} = [0, 1]$, it follows immediately that $\widehat{f}_h(\mathbf{x}^Q, t) \le f(\mathbf{x}^Q, t)$.

Case 2: $N_h(\mathbf{x}, p) > 0$. In this case, the pessimistic estimator evaluates to

$$\tilde{f}_h(\mathbf{x}^Q, t) = \widehat{f}_h(\mathbf{x}^Q, t) - \sqrt{\frac{\log n}{N_h(\mathbf{x}^Q, t)}} - h^\beta,$$

and by Proposition 6, with probability at least $1 - 1/n$, the error is bounded as

$$\big|f(\mathbf{x}^Q, t) - \widehat{f}_h(\mathbf{x}^Q, t)\big| \le h^\beta + \sqrt{\frac{\log n}{N_h(\mathbf{x}^Q, t)}}.$$

Therefore, we have

$$f(\mathbf{x}^Q, t) \ge \widehat{f}_h(\mathbf{x}^Q, t) - h^\beta - \sqrt{\frac{\log n}{N_h(\mathbf{x}^Q, t)}} = \tilde{f}_h(\mathbf{x}^Q, t).$$

□

C.1 Proof of Proposition 6

Proof. Given a target context $\mathbf{x}^Q$ and treatment candidate $t \in \mathcal{T}$, we have

$$\begin{aligned}
&\big|\widehat{f}_h(\mathbf{x}^Q, t) - f(\mathbf{x}^Q, t)\big| \\
&\quad = \left| \frac{\sum_{i=1}^n Y_i I_{h,i}(\mathbf{x}^Q, t)}{N_h(\mathbf{x}^Q, t) \vee 1} - f(\mathbf{x}^Q, t) \right| \\
&\quad = \left| 0 \cdot \mathbb{1}\{N_h(\mathbf{x}^Q, t) = 0\} + \mathbb{1}\{N_h(\mathbf{x}^Q, t) > 0\} \frac{\sum_{i=1}^n Y_i I_{h,i}(\mathbf{x}^Q, t)}{N_h(\mathbf{x}^Q, t)} - f(\mathbf{x}^Q, t) \right| \\
&\quad = \left| \mathbb{1}\{N_h(\mathbf{x}^Q, t) = 0\}\big(0 - f(\mathbf{x}^Q, t)\big) + \mathbb{1}\{N_h(\mathbf{x}^Q, t) > 0\} \left(\frac{\sum_{i=1}^n Y_i I_{h,i}(\mathbf{x}^Q, t)}{N_h(\mathbf{x}^Q, t)} - f(\mathbf{x}^Q, t) \right) \right| \\
&\quad \le \underbrace{\left| \mathbb{1}\{N_h(\mathbf{x}^Q, t) = 0\}\big(0 - f(\mathbf{x}^Q, t)\big) \right|}_{E_1} + \underbrace{\left| \mathbb{1}\{N_h(\mathbf{x}^Q, t) > 0\} \left(\frac{\sum_{i=1}^n Y_i I_{h,i}(\mathbf{x}^Q, t)}{N_h(\mathbf{x}^Q, t)} - f(\mathbf{x}^Q, t) \right) \right|}_{E_2},
\end{aligned}$$

where we recall that $I_{h,i}(\mathbf{x}^Q, t) = \mathbb{1}\big\{(\|\mathbf{X}_i^P - \mathbf{x}^Q\|_\infty \vee |t_i^P - t|) \le h\big\}$. We bound E_1 and E_2 in order.

Bounding E_1. Since $\|f\|_\infty \le 1$, we have

$$E_1 = \big|\mathbb{1}\{N_h(\mathbf{x}^Q, t) = 0\}\big(0 - f(\mathbf{x}^Q, t)\big)\big| = \mathbb{1}\{N_h(\mathbf{x}^Q, t) = 0\}|f(\mathbf{x}^Q, t)| \le \mathbb{1}\{N_h(\mathbf{x}^Q, t) = 0\}.$$

Bounding E_2. For notational simplicity, from now on we omit $\mathbb{1}\{N_h(\mathbf{x}^Q, t) > 0\}$ multiplied in every term. Using the triangle inequality, we decompose E_2 into two terms:

$$E_2 \le \Bigg(\underbrace{\big|f(\mathbf{x}^Q, t) - \bar{f}\big|}_{E_{21}} + \underbrace{\left|\bar{f} - \frac{\sum_{i=1}^n Y_i I_{h,i}(\mathbf{x}^Q, t)}{N_h(\mathbf{x}^Q, t)}\right|}_{E_{22}}\Bigg),$$

where $\bar{f}$ is defined as

$$\bar{f}(\mathbf{x}^Q, t) = \frac{1}{N_h(\mathbf{x}^Q, t)} \sum_{i=1}^n f(\mathbf{X}_i^P, t_i^P) I_{h,i}(\mathbf{x}^Q, t).$$

Using the following algebraic equality:

$$f(\mathbf{x}^Q, t) = \frac{1}{N_h(\mathbf{x}^Q, t)} \sum_{i=1}^n f(\mathbf{x}^Q, t) I_{h,i}(\mathbf{x}^Q, t), \tag{14}$$

we can bound E_{21} as follows:

$$\begin{aligned}
E_{21} &= \big|f(\mathbf{x}^Q, t) - \bar{f}(\mathbf{x}^Q, t)\big| \\
&\overset{(i)}{\le} \frac{1}{N_h(\mathbf{x}^Q, t)} \sum_{i=1}^n \big|f(\mathbf{x}^Q, t) - f(\mathbf{X}_i^P, t_i^P)\big| \, I_{h,i}(\mathbf{x}^Q, t) \\
&\overset{(ii)}{\le} \frac{1}{N_h(\mathbf{x}^Q, t)} \sum_{i=1}^n L\left(\|\mathbf{X}_i^P - \mathbf{x}^Q\|_\infty \vee |t_i^P - p|\right)^\beta I_{h,i}(\mathbf{x}^Q, t) \\
&\overset{(iii)}{\le} \frac{1}{N_h(\mathbf{x}^Q, t)} \sum_{i=1}^n L h^\beta I_{h,i}(\mathbf{x}^Q, t) \\
&= \frac{1}{N_h(\mathbf{x}^Q, t)} L h^\beta N_h(\mathbf{x}^Q, t) = L h^\beta,
\end{aligned}$$

where step (i) uses (14), step (ii) applies Assumption 1, and step (iii) uses the definition $I_{h,i}(\mathbf{x}^Q, t) = \mathbb{1}\big\{(\|\mathbf{X}_i - \mathbf{x}^Q\|_\infty \vee |p_i - p|) \le h\big\}$.

Next, we bound E_{22}. Using the following algebraic equality:

$$Y_i^P = \frac{1}{N_h(\mathbf{x}^Q, t)} \sum_{i=1}^n Y_i^P I_{h,i}(\mathbf{x}^Q, t) \le h\big\},$$

we have

$$E_{22} = \left|\frac{1}{N_h(\mathbf{x}^Q, t)} \sum_{i=1}^n \big(Y_i^P - f(\mathbf{X}_i, p_i)\big) I_{h,i}(\mathbf{x}^Q, t)\right|.$$

By the problem setup given in Section 1, conditional on $\{(\mathbf{X}_i^P, t_i^P)\}_{i=1}^n$, $\big(Y_i^P - f(\mathbf{X}_i^P, t_i^P)\big)$'s are zero-mean independent 1-sub-Gaussians, and thus E_2 is the absolute value of the average of $N_h(\mathbf{x}^Q, t)$ zero-mean independent 1-sub-Gaussians. By the sub-Gaussian concentration, with probability at least $1 - 1/n$, we have

$$E_{22} \le \sqrt{\frac{2 \log 2n}{N_h(\mathbf{x}^Q, t)}}.$$

Conclusion. Collecting the bounds for E_1, E_{21} and E_{22}, with probability at least $1 - 1/n$, we have

$$\big|\widehat{f}_h(\mathbf{x}^Q, t) - f(\mathbf{x}^Q, t)\big| \le \mathbb{1}\{N_h(\mathbf{x}^Q, t) = 0\} + \mathbb{1}\{N_h(\mathbf{x}^Q, t) > 0\}\left(L h^\beta + \sqrt{\frac{2 \log 2n}{N_h(\mathbf{x}^Q, t)}}\right).$$

This completes the proof of Proposition 6. □

D Proof of Theorem 2

D.1 Proof of Upper Bound

This section proves the equation (6) in Theorem 2.

Proof. The best policy $\pi^* \in \arg\max_\pi V(\pi)$ maps each context $\mathbf{x}$ to an optimal treatment $t^*(\mathbf{x}) \in \arg\max_{t\in\mathcal{T}} f(\mathbf{x}, t)$, with ties resolved by a fixed measurable selection t^*. For a given target context $\mathbf{x}^Q \in \mathcal{X}$, recall that we denote the single optimal treatment that satisfies Assumption 4 as $t^\dagger(\mathbf{x})$. Conditioned on the event from Proposition 6, the suboptimality of the NW-LCB policy relative to any optimal treatment $t^*(\mathbf{x})$ is bounded by a constant multiple of the lower confidence interval's length. This is shown by the following series of inequalities, that holds with probability at least $1 - 1/n$:

$$\begin{aligned}
f\big(\mathbf{x}^Q, t^*(\mathbf{x}^Q)\big) - f\big(\mathbf{x}^Q, \hat{\pi}(\mathbf{x}^Q)\big) &\overset{(i)}{\le} f\big(\mathbf{x}, t^*(\mathbf{x})\big) - \tilde{f}\big(\mathbf{x}, \hat{\pi}(\mathbf{x})\big) \\
&\overset{(ii)}{=} f\big(\mathbf{x}, t^\dagger(\mathbf{x})\big) - \tilde{f}\big(\mathbf{x}, \hat{\pi}(\mathbf{x})\big) \\
&\overset{(iii)}{\le} f\big(\mathbf{x}, t^\dagger(\mathbf{x})\big) - \tilde{f}\big(\mathbf{x}, t^\dagger(\mathbf{x})\big) \\
&\overset{(iv)}{\lesssim} h^\beta \mathbb{1}\{N_h\big(\mathbf{x}, t^\dagger(\mathbf{x})\big) > 0\} + \log n \frac{\mathbb{1}\{N_h\big(\mathbf{x}, t^\dagger(\mathbf{x})\big) > 0\}}{\sqrt{N_h\big(\mathbf{x}, t^\dagger(\mathbf{x})\big)}},
\end{aligned}$$

where step (i) Lemma 1 which holds with high probability, step (ii) uses the fact that both of $t^*(\mathbf{x})$ and $t^\dagger(\mathbf{x})$ belong to $\arg\max_{t\in\mathcal{T}} f(\mathbf{x}^Q, p^Q)$, step (iii) uses the definition of NW-LCB policy as the maximizer of $\tilde{f}(\mathbf{x}, p)$ for a fixed $\mathbf{x}$, presented in (3), and step (iv) uses Proposition 6 at point $\big(\mathbf{x}, t^\dagger(\mathbf{x})\big)$.

To control the suboptimality, we must control the expected length of the lower confidence interval, taking into account the covariate shift:

$$\begin{aligned}
&V(\pi^*) - V(\hat{\pi}) = \\
&\quad \mathbb{E}\left[f\big(\mathbf{X}^Q, t^\dagger(\mathbf{X}^Q)\big) - f\big(\mathbf{X}^Q, \hat{\pi}(\mathbf{X}^Q)\big)\right], \\
&\quad \lesssim h^\beta \mathbb{E}\left[\mathbb{1}\{N_h\big(\mathbf{X}^Q, t^\dagger(\mathbf{X}^Q)\big) > 0\}\right] + \log n\, \mathbb{E}\Bigg[\frac{\mathbb{1}\{N_h\big(\mathbf{X}^Q), t^\dagger(\mathbf{X}^Q)\big) > 0\}}{\sqrt{N_h\big(\mathbf{X}^Q), t^\dagger(\mathbf{X}^Q)\big)}}\Bigg] \\
&\quad \le h^\beta + \log n\, \mathbb{E}\Bigg[\frac{\mathbb{1}\{N_h\big(\mathbf{X}^Q), t^\dagger(\mathbf{X}^Q)\big) > 0\}}{\sqrt{N_h\big(\mathbf{X}^Q), t^\dagger(\mathbf{X}^Q)\big)}}\Bigg] \qquad (15)
\end{aligned}$$

The following lemma bounds the expectation in the last display.

Lemma 7. *With probability at least $1 - 1/n$, we have*

$$\mathbb{E}\Bigg[\frac{\mathbb{1}\{N_h\big(\mathbf{X}^Q), t^\dagger(\mathbf{X}^Q)\big) > 0\}}{\sqrt{N_h\big(\mathbf{X}^Q), t^\dagger(\mathbf{X}^Q)\big)}}\Bigg] \lesssim (\zeta\, n\, h)^{-1/2} h^{-(d+\alpha)/2}.$$

Proof of Lemma 7 is provided in Appendix D.6.1. In Lemma 7, the expecation is taken with respect to two sources of randomness the source context-treatment pair distribution μ, which is used to construct the prediction interval, and the target context distribution Q, which governs the new contexts for which we make a prediction.

Combining (15), (39), and Lemma 7, we have:

$$V(\pi^*) - V(\hat{\pi}) \lesssim h^\beta + (n\, \zeta\, h)^{-1/2} h^{-(d+\alpha)/2}$$

hiding the log factor. If we set

$$h = \Theta\left((\zeta n)^{-\frac{1}{2\beta+\alpha+d+1}}\right),$$

and

$$\bar{h} = \Omega\left((\zeta n)^{-\frac{1}{2\beta+\alpha+d+1}}\right),$$

we have

$$\begin{aligned} V(\pi^*) - V(\hat{\pi}) &\lesssim (\zeta n)^{-\frac{\beta}{2\beta+\alpha+d+1}} + (\zeta\, n)^{-1/2}(\zeta n)^{\frac{d+1+\alpha}{2}\frac{1}{2\beta+\alpha+d+1}} \\ &= (\zeta n)^{-\frac{\beta}{2\beta+\alpha+d+1}} + (\zeta\, n)^{-\frac{1}{2}(1-\frac{d+1+\alpha}{2\beta+\alpha+d+1})} \\ &\lesssim (\zeta n)^{-\frac{\beta}{2\beta+\alpha+d+1}}. \end{aligned}$$

This completes the proof of equation (6) of Theorem 2. □

D.2 Proof of Lower bound: Outline

This section and subsequent sections prove the equation (7) in Theorem 2. We use the following version of Fano's Lemma which does not require a metric.

Theorem 8 (Fano's Lemma)**.** *Let Θ be a class of distributions. Consider a loss function $L : \Theta \times [0,1] \to \mathbb{R}_+$, which evaluates the quality of a treatment for a given distribution. If we have $\theta_1, \ldots, \theta_m \in \Theta$, such that*

$$L(\theta_i, p) + L(\theta_j, p) \geq \Delta, \quad \forall i \neq j \in [m], t \in \mathcal{T},$$

we have

$$\inf_{\pi} \sup_{\theta \in \Theta} \mathbb{E}_{D \sim \theta}[L(\theta, \pi(D))] \geq \frac{\Delta}{2} \inf_{\Psi} \frac{1}{m} \sum_{i=1}^{m} \theta_i(\Psi \neq i) \geq \frac{\Delta}{2}\left\{1 - \frac{1}{\log m}\left(\frac{1}{m^2}\sum_{i,j=1}^{m} \mathrm{KL}(\theta_i \| \theta_j) + \log 2\right)\right\}$$

where the infimum on the first term is taken over all pricing policies π, while the infimum in the second term is taken over all measurable tests $\Psi : \mathcal{X} \to [m]$.

The proof of Lemma 8 is provided in Appendix D.6.2.

The proof follows the following intermediate steps:

1. Construction of problem instances,
2. Application of Fano's lemma.

D.3 Construction of problem instances

D.3.1 Construction of domain grids and packing indices

When constructing the distributions and the reward function, we discretize the domains $\mathcal{X}$ and $\mathcal{T}$ using a grid with a spacing determined by a radius parameter, $r \in (0, 1/2)$. This parameter will be specified later. The same radius r is also used to construct the index set for the packing, which is a finite, well-separated collection of distributions satisfying the condition of Lemma 8.

Context grid. We partition the context space $\mathcal{X} = [0,1]^d$ into a collection of hypercubes, each with a side length of $2r$. Let $\mathcal{S}_{\mathcal{X},r}$ denote the set of centers for these hypercubes. The coordinates of these grid points originate from r and increment with a uniform spacing of $2r$. The cardinality is $\lfloor 1/(2r) \rfloor^d$. More formally:

$$\mathcal{S}_{\mathcal{X},r} := \left\{\mathbf{x} = (x_1, x_2, \ldots, x_d) \in [0,1]^d \mid x_i = \left(k_i - \tfrac{1}{2}\right) 2r,\ k_i \in \left[\left\lfloor \tfrac{1}{2r} \right\rfloor\right],\ \forall i \in [d]\right\} = \left\{\mathbf{x}_1^*, \ldots, \mathbf{x}_{\lfloor 1/(2r) \rfloor^d}^*\right\}. \tag{16}$$

treatment grid. We partition the treatment space $\mathcal{T} = [0,1]$ into intervals, each with a length of r. Let $\mathcal{S}_{\mathcal{T},r}$ represent the centers of these intervals. These grid points begin at $\frac{1}{2}r$ and increase by increments of r. The cardinality is $\lfloor 1/r \rfloor$. More formally:

$$\mathcal{S}_{\mathcal{T},r} = \left\{(k - \tfrac{1}{2})r \ :\ k \in \left[\left\lfloor \tfrac{1}{r} \right\rfloor\right]\right\} = \left\{t_1^*, \ldots, p_{\lfloor 1/r \rfloor}^*\right\}. \tag{17}$$

Packing indices. We now construct an index set that will parametrize the hard instances used in our Fano-type argument. Following a standard technique in nonparametric statistics (see, for example, Example 15.15 in Wainwright [27] for a detailed exposition), we construct a local packing of the function space (see Appendix D.3.3 for details). First, define the integer

$$m := \lfloor c_m/r \rfloor^d \tag{18}$$

for a constant $0 < c_m < 1/2$. Using the Varshamov-Gilbert bound (see, e.g., Example 5.3 of Wainwright 27) , we can find a set of well-separated vectors

$$\Omega_m := \{\boldsymbol{\omega}_i\}_{i=0}^M \subset \{\pm 1\}^m \tag{19}$$

with sufficiently large cardinality and separation:

$$\log_2(M) \geq \frac{m}{8} \quad \text{and} \quad \rho(\boldsymbol{\omega}_i, \boldsymbol{\omega}_j) \geq \frac{m}{8}, \quad \forall 0 \leq i < j \leq M, \tag{20}$$

where $\rho(\boldsymbol{\omega}, \boldsymbol{\omega}')$ is the Hamming distance, which measures the number of indices at which the vectors $\boldsymbol{\omega}$ and $\boldsymbol{\omega}'$ differ.

D.3.2 Construction of problem instances - context and treatment

Using the grids and packing index set constructed in Appendix D.3.1, we define a collection of distributions:

$$\mathcal{H}_{\Omega_m} := \{(Q, \mu, f_{\boldsymbol{\omega}}) \mid \boldsymbol{\omega} \in \Omega_m\}\,. \tag{21}$$

For simplicity, only the reward function $f_{\boldsymbol{\omega}}$ depends on the binary vector $\boldsymbol{\omega} \in \Omega_m$, while the target context distribution Q and source context-treatment pair distribution μ remain fixed across all instances. We first construct Q and μ, and define $f_{\boldsymbol{\omega}}$ in Appendix D.3.3.

Target context distribution. We define the target context distribution Q by specifying its piecewise constant density, denoted $q(\mathbf{x})$. This density is constructed using small ℓ_∞ balls of radius $r/4$ and large ℓ_∞ balls of radius r, both centered at the context grid points $\mathcal{S}_{\mathcal{X},r}$ (as defined in Equation (16)). Formally, for any $\mathbf{x} \in [0,1]^d$:

$$q(\mathbf{x}) := \begin{cases} q_1, & \text{if } \mathbf{x} \in \bigcup_{i=1}^m B_{r/4}(\mathbf{x}_i^*), \\ q_0, & \text{if } \mathbf{x} \in \mathcal{X} \setminus \bigcup_{i=1}^m B_r(\mathbf{x}_i^*), \\ 0, & \text{otherwise}, \end{cases} \tag{22}$$

where

$$q_1 := \frac{r^d}{\mathrm{Leb}\big(B_{r/4}(\mathbf{x}_1^*)\big)} \quad \text{and} \quad q_0 := \frac{1 - mr^d}{\mathrm{Leb}\big(\mathcal{X} \setminus \bigcup_{i=1}^m B_r(\mathbf{x}_i^*)\big)}.$$

Source context-treatment pair distribution. We define the source context-treatment pair distribution μ by first specifying the source context distribution P and then the conditional source treatment distribution π_{off}. The source context distribution P is defined by its piecewise constant density, $p(\mathbf{x})$,which is a modification of $q(\mathbf{x})$:

$$p(\mathbf{x}) := \begin{cases} C_\alpha r^\alpha q_1, & \text{if } \mathbf{x} \in \bigcup_{i=1}^m B_{r/4}(\mathbf{x}_i^*), \\ \delta, & \text{if } \mathbf{x} \in \bigcup_{i=1}^m B_r(\mathbf{x}_i^*) \setminus B_{r/2}(\mathbf{x}_i^*), \\ q_0, & \text{if } \mathbf{x} \in \mathcal{X} \setminus \bigcup_{i=1}^m B_r(\mathbf{x}_i^*), \\ 0, & \text{otherwise}, \end{cases} \tag{23}$$

where the constant δ is chosen to ensure that $p(\mathbf{x})$ integrates to one:

$$\delta := \frac{1 - c_\alpha r^\alpha q_1 \,\mathrm{Leb}\big(\bigcup_{i=1}^m B_{r/4}(\mathbf{x}_i^*)\big) - q_0 \,\mathrm{Leb}\big(\mathcal{X} \setminus \bigcup_{i=1}^m B_r(\mathbf{x}_i^*)\big)}{\mathrm{Leb}\big(\bigcup_{i=1}^m B_r(\mathbf{x}_i^*) \setminus B_{r/2}(\mathbf{x}_i^*)\big)},$$

and α and C_α are the transfer exponent and its corresponding constant, as defined in Assumption 3. Then, we define the source treatment distribution given the source context $\mathbf{x}$, denoted $\pi_{\text{off}}(\cdot \mid \mathbf{x})$, as a binary pricing policy that does not depend on the context. Its conditional density $\wp(p \,|\mathbf{x})$ is formally defined as:

$$\wp(p \mid \mathbf{x}) := \begin{cases} \zeta, & \text{if } p \in [\tilde{p} - r/2, \tilde{p} + r/2], \\ \dfrac{1 - r\zeta}{1 - r}, & \text{otherwise}, \end{cases} \tag{24}$$

where $\tilde{p} \in \mathcal{S}_{P,r}$ be defined later. where we recall that ζ is the exploration coefficient defined in .

The following two lemmas demonstrate that our constructions satisfy the assumptions for the transfer exponent (Assumption 3) and the exploration coefficient (Assumption 4).

Lemma 9 (Transfer exponent)**.** *For the source covariate distribution constructed as in* (23) *and the target covariate distribution Q constructed as in* (22)*, the transfer exponent of P with respect to Q is α, with corresponding constant C_α.*

Proof of Lemma 9 is provided in Appendix D.6.3.

Lemma 10 (Near-optimal treatment exploration)**.** *The data collecting policy π_{off} with conditional density $\wp(p \mid \mathbf{x})$ defined as* (24) *satisfies the near-optimal treatment exploration assumption with coefficient ζ.*

Proof of Lemma 10 is provided in Appendix D.6.4.

D.3.3 Constructing the reward distributions

For both target and source data, let the random reward, conditional on the context $\mathbf{x}$ and treatment p, be a Bernoulli random variable with success probability $f_{\boldsymbol{\omega}}(\mathbf{x}, t)$. This setting satisfies the 1-sub-Gaussian assumption. The reward functions $\{f_{\boldsymbol{\omega}}\}_{\boldsymbol{\omega} \in \Omega_m}$ are constructed as follows. We begin by defining a simple Hölder continuous function.

Lemma 11. *The function $\phi : \mathbb{R}_+ \to [0,1]$ defined by*

$$\phi(z) := \begin{cases} 1, & \text{if } 0 \le z < \frac{1}{4}, \\ (2-4z)^\beta, & \text{if } \frac{1}{4} \le z < \frac{1}{2}, \\ 0, & \text{otherwise}, \end{cases} \tag{25}$$

is Hölder continuous with exponent $\beta \in (0,1]$ and Hölder constant 4^β.

Proof of Lemma 11 is provided in Appendix D.6.5.

Next, define the function $\varphi_\beta : [0,1]^d \times \mathcal{T} \to [0, 1/4]$ via

$$\varphi_\beta(\mathbf{x}, t) := C_\beta \, r^\beta \, \phi\left(\frac{1}{r} \max\{\|\mathbf{x}\|_\infty, |p|\}\right), \tag{26}$$

where $C_\beta := \min\{C_{\mathrm{H}}, 1\}/4^\beta$.

We add this φ_β function as bumps, centered at the grid points $S_{\mathcal{X},r}$ in the context space, and also at a specific point t_u^* within $S_{\mathcal{T},r}$ in the treatment space. This point t_u^* is defined as follows. Fix a policy π. For any grid point $p_i^* \in \mathcal{S}_{\mathcal{T},r}$, define a indicator variable

$$I^Q_{r,\pi}(p_i^*) := \mathbb{1}\{\pi(\mathbf{X}^Q) \in [t_i^* - r/2, t_i^* + r/2]\}. \tag{27}$$

For any policy π, we have

$$\sum_{i=1}^{\lfloor 1/r \rfloor} \mathbb{E}_Q[I^Q_{r,\pi}(p_i^*)] = \mathbb{E}_Q\left[\sum_{i=1}^{\lfloor 1/r \rfloor} I^Q_{r,\pi}(p_i^*)\right] = 1.$$

Therefore, by the pidgeon's principle, there must exist at least one grid point $t_u^* \in \mathcal{S}_{\mathcal{T},r}$ such that

$$\mathbb{E}_Q[I^Q_{r,\pi}(p_u^*)] \le \frac{1}{\lfloor 1/r \rfloor}. \tag{28}$$

We denote $\tilde{p} := t_u^*$. With this definition, we finally construct a Hölder continuous reward function:

Lemma 12 (Hölder continuous reward)**.** *Fix β $in(0,1]$. for any $\boldsymbol{\omega} \in \Omega_m$, the reward function $f_{\boldsymbol{\omega}} : \mathcal{X} \times \mathcal{T} \to \mathcal{Y}$ defined as*

$$f_{\boldsymbol{\omega}}(\mathbf{x}, t) := \frac{1}{2} + \sum_{i=1}^{m} \omega_i \, \varphi_\beta\big((\mathbf{x} - \mathbf{x}_i^*),\, (p - \tilde{p})\big) \cdot \mathbb{1}\left\{\mathbf{x} \in B_r(\mathbf{x}_i^*)\right\}. \tag{29}$$

is Hölder continuous with exponent β and Hölder constant $C_\beta > 0$.

The proof of Lemma 12 is provided in Appendix D.6.6.

D.4 Application of the Fano's method

Using the problem instances constructed in Appendix D.3, we apply Lemma 8 to derive the minimax lower bound. Appendix D.4.1 demonstrates that these instances are sufficiently separated in the parameter space by providing a lower bound on the sub-optimality gap.

D.4.1 Lower bounding the sub-optimal gap

The functions ϕ, φ_β, and $f_{\boldsymbol{\omega}}$, defined in Equations (25), (26), and (29) respectively, are designed to satisfy the lemmas 13, 14, 15, 17 and Corollary 16, which will lower bounding the sub-optimal gap. We begin with Lemma 13, which states that the constructed reward function is not entirely composed of bumps, but features flat regions in both context and treatment when they are $r/2$ away from the grid points used for its definition.

Lemma 13 (Flat regions)**.** *For any $\boldsymbol{\omega} \in \Omega_m$, the reward function $f_{\boldsymbol{\omega}}(\mathbf{x}, t)$ defined in* (29) *has a value of* $1/2$ *under the following conditions:*

(i) For any $\mathbf{x} \in \mathcal{X} \setminus \bigcup_{i=1}^m B_{r/2}(\mathbf{x}_i^*)$ and any $p \in [0, 1]$, we have $f_{\boldsymbol{\omega}}(\mathbf{x}, t) = \frac{1}{2}$.

(ii) For any $p \in [0, 1] \setminus [\tilde{p} - r/2, \tilde{p} + r/2]$ and any $\mathbf{x} \in \mathcal{X}$, we have $f_{\boldsymbol{\omega}}(\mathbf{x}, t) = \frac{1}{2}$.

The proof of Lemma 13 is omitted because it follows directly from the fact that $\phi(z)$, defined in (25), vanishes for $z > 1/2$, and that $\varphi_\beta(\mathbf{x}, t)$, defined in (26), is constructed using the max norm.

Next, the following lemma demonstrates the existence of the minimum optimal treatment.

Lemma 14. *For any $\boldsymbol{\omega} \in \Omega_m$, the minimum optimal treatment, denoted as*

$$p_{\boldsymbol{\omega}}^\dagger(\mathbf{x}) := \min \left\{ t' \in \arg\max_{p \in [0,1]} f_{\boldsymbol{\omega}}(\mathbf{x}, t) \right\}, \tag{30}$$

is well-defined.

The proof of Lemma 14 is provided in Appendix D.6.7.

Next, the following lemma states that when the context is sufficiently close to one of the grid points, the minimum optimal treatments corresponding to different bump directions are well-separated.

Lemma 15 (optimal treatment separation)**.** *Given $\boldsymbol{\omega} \in \Omega_m$ and $\mathbf{x} \in \mathcal{X}$, if there exists $i \in [m]$ such that $\mathbf{x} \in B_{r/4}(\mathbf{x}_i^*)$:*

1. *If $\omega_i = -1$, then $p_{\boldsymbol{\omega}}^\dagger(\mathbf{x}) = 0$.*
2. *If $\omega_i = 1$, then $p_{\boldsymbol{\omega}}^\dagger(\mathbf{x}) = \tilde{p} - r/4$.*

Proof of Lemma 15 is provided in Appendix D.6.8.

For the next lemma, we define the function $h : \mathcal{T} \to \{0, 1\}$ to characterize indicate whether the treatment p falls outside the flat region.

$$h(p) := \mathbb{1}\left\{ p \in \left(\tilde{p} - \frac{r}{2}, \tilde{p} + \frac{r}{2}\right] \right\}.$$

The following is a corollary of Lemma 15.

Corollary 16. *Fix $\mathbf{x} \in \mathcal{X}$ and two distinct $\boldsymbol{\omega}, \boldsymbol{\omega}' \in \Omega_m$. If there exists an index $i \in [m]$ such that $\omega_i \neq \omega_i'$ and $\mathbf{x} \in B_{r/4}(\mathbf{x}_i^*)$, then exactly one of $p_{\boldsymbol{\omega}}^\dagger(\mathbf{x})$ and $t_{\boldsymbol{\omega}'}^*(\mathbf{x})$ lies in the flat region presented in Lemma 13, while the other lies outside it. More formally,*

$$h(p_{\boldsymbol{\omega}}^\dagger(\mathbf{x})) \neq h(t_{\boldsymbol{\omega}'}^*(\mathbf{x})).$$

The next lemma states that for contexts and treatments sufficiently close to grid points, the sub-optimality gap directly reflects whether the optimal treatment and the chosen treatment p fall into different regions (either flat or bump regions).

Lemma 17. *For any $\boldsymbol{\omega} \in \Omega_m$, $\mathbf{x} \in \bigcup_{i=1}^m B_{r/4}(\mathbf{x}_i^*)$, and $p \in \bigcup_{i=1}^{\lfloor 1/r \rfloor} [p_i^* - r/4, p_i^* + r/4]$, the following holds:*

$$f_{\boldsymbol{\omega}}(\mathbf{x}, p_\omega^*(\mathbf{x})) - f_{\boldsymbol{\omega}}(\mathbf{x}, t) = C_\beta r^\beta \mathbb{1}\{h(p_\omega^*(\mathbf{x})) \neq h(p)\}.$$

The proof of Lemma 17 is provided in Appendix D.6.9.

To use Lemma 17, it will be useful if we can only consider policies that only assignes treatments $r/4$-close to the gridpoints $S_{\mathcal{T},r}$, defined in (17).

Lemma 18. *For any treatment policy π, for any $\mathbf{x} \in \mathcal{X}$, if*

$$\pi(\mathbf{x}) \in \bigcup_{i=1}^{\lfloor 1/r \rfloor} \left([p_i^* - r/2, p_i^* + r/2] \setminus [p_i^* - r/4, p_i^* + r/4] \right), \tag{31}$$

then there exists a policy π' with

$$\pi'(\mathbf{x}) \in \bigcup_{i=1}^{\lfloor 1/r \rfloor} [p_i^* - r/4, p_i^* + r/4],$$

and satisfies

$$f_{\boldsymbol{\omega}}(\mathbf{x}, \pi'(\mathbf{x})) \geq f_{\boldsymbol{\omega}}(\mathbf{x}, \pi(\mathbf{x}))$$

Proof of Lemma 18 is provided in Appendix D.6.10.

Conclusion. By Lemma 18, when lower bounding the sub-optimal gap, it suffices to consider treatment policies where

$$\pi(\mathbf{x}) \in \bigcup_{i=1}^{\lfloor 1/r \rfloor} [p_i^* - r/4, p_i^* + r/4] \text{ for all } \mathbf{x} \in \mathcal{X}. \tag{32}$$

For any policy π satisfying (32) and for any $\boldsymbol{\omega} \in \Omega_m$, let us define

$$\mathrm{SubOpt}(\pi, Q, \boldsymbol{\omega}) := \mathbb{E}_{\boldsymbol{\omega}} \left[f_{\boldsymbol{\omega}}(\mathbf{X}^Q, p_{\boldsymbol{\omega}}^\dagger(\mathbf{X}^Q)) - f_{\boldsymbol{\omega}}(\mathbf{X}^Q, \pi(\mathbf{X}^Q)) \right], \tag{33}$$

where $E_{\boldsymbol{\omega}}$ denotes the expectation under the distribution $(Q, \mu, f_{\boldsymbol{\omega}})$ defined in Appendices D.3.2 and D.3.3. By Lemma 17, we have that

$$\begin{aligned}
\mathrm{SubOpt}(\pi, Q, \boldsymbol{\omega}) &= C_\beta r^\beta \mathbb{E}_{\boldsymbol{\omega}} \left[\mathbb{1}\{h(p_{\boldsymbol{\omega}}^*(\mathbf{X}^Q)) \neq h(\pi(\mathbf{X}^Q))\} \mathbb{1}\{\mathbf{X}^Q \in \bigcup_{i=1}^m B_{r/4}(x_i^*)\} \right] \\
&= C_\beta r^\beta \sum_{i=1}^m \mathbb{E}_{\boldsymbol{\omega}} \left[\mathbb{1}\{h(p_{\boldsymbol{\omega}}^*(\mathbf{X}^Q)) \neq h(\pi(\mathbf{X}^Q))\} \mathbb{1}\{\mathbf{X}^Q \in B_{r/4}(x_i^*)\} \right].
\end{aligned}$$

Therefore, for any policy π satisfying (32) and for any $\boldsymbol{\omega} \neq \boldsymbol{\omega}' \in \Omega_m$, if we set $r = (\zeta n)^{-1/(d+1+2\beta+\alpha)}$, then we have

$$\begin{aligned}
&\mathrm{SubOpt}(\pi, Q, \boldsymbol{\omega}) + \mathrm{SubOpt}(\pi, Q, \boldsymbol{\omega}') \\
&= C_\beta r^\beta \sum_{i=1}^m \mathbb{E}_{\boldsymbol{\omega}} \left[\left(\mathbb{1}\{h(p_{\boldsymbol{\omega}}^*(\mathbf{X}^Q)) \neq h(\pi(\mathbf{X}^Q))\} + \mathbb{1}\{h(p_{\boldsymbol{\omega}'}^*(\mathbf{X}^Q)) \neq h(\pi(\mathbf{X}^Q))\} \right) \mathbb{1}\{\mathbf{X}^Q \in B_{r/4}(x_i^*)\} \right] \\
&\geq C_\beta r^\beta \sum_{i=1}^m \mathbb{E}_{\boldsymbol{\omega}} \left[\mathbb{1}\{h(p_{\boldsymbol{\omega}}^*(\mathbf{X}^Q)) \neq h(p_{\boldsymbol{\omega}'}^*(\mathbf{X}^Q))\} \mathbb{1}\{\mathbf{X}^Q \in B_{r/4}(x_i^*)\} \right] \\
&\overset{(i)}{\geq} C_\beta r^\beta \sum_{i=1}^m \mathbb{E}_{\boldsymbol{\omega}} \left[\mathbb{1}\{\omega_i \neq \omega_i'\} \mathbb{1}\{\mathbf{X}^Q \in B_{r/4}(x_i^*)\} \right] \\
&= C_\beta r^{d+\beta} \rho(\boldsymbol{\omega}, \boldsymbol{\omega}') \\
&\overset{(iii)}{\geq} C_\beta r^{d+\beta} \frac{m}{8} \\
&\overset{(iv)}{\geq} C_\beta r^{d+\beta} \lfloor c_m / r \rfloor^d \\
&\geq C_\beta c_m r^\beta \\
&\gtrsim r^\beta \\
&\overset{(v)}{\gtrsim} (\zeta n)^{-\beta/(d+1+2\beta+\alpha)}.
\end{aligned} \tag{34}$$

where step (i) uses Corollary 16, step (ii) uses the piecewise construction of target context distribution in (22), step (iii) uses the cardinality of the packing in (20), step (iv) uses the definition of m given in (18), and step (v) uses $r = (\zeta n)^{-1/(d+1+2\beta+\alpha)}$.

D.5 Upper bounding the distribution distance

For each problem instance indexed by $\boldsymbol{\omega} \in \Omega_m$, its full joint distribution is denoted by $\theta_{\boldsymbol{\omega}}$, which we formally define as follows.

Definition 2. *Fix a policy π. Denote the source data as $\mathcal{D}^P := \{\mathbf{X}_t^P, p_t^P, Y_t^P\}_{t=1}^n$ and the target data as $\mathcal{D}^Q := \{\mathbf{X}^Q, \pi(\mathbf{X}^Q), Y^Q\}$, which represents a single time horizon. For each $\boldsymbol{\omega} \in \Omega_m$, let $\theta_{\boldsymbol{\omega},\pi}$ be the joint distribution of the random variables in $\mathcal{D}^P$ and $\mathcal{D}^Q$, which is induced by μ, Q, $f_{\boldsymbol{\omega}}$ and π.*

The data generating process under $\theta_{\boldsymbol{\omega},\pi}$, and associated notations, are defined as follows:

- The source data $\mathcal{D}^P$ is generated as follows. A source context $\mathbf{X}^P$ is drawn from the source context distribution P, with density $p(\cdot)$. The source treatment p_t^P is drawn with respect to the conditional density $\wp(\cdot \mid \mathbf{X}^P)$. The reward is drawn from $\text{Bernoulli}\big(f_{\boldsymbol{\omega}}(\mathbf{X}_t^P, p_t^P)\big)$, with likelihood denoted as $\vartheta_{f_{\boldsymbol{\omega}}}(\cdot\,; \mathbf{x}, \mathbf{p})$.
- The target data $\mathcal{D}^Q$ is generated as follows. A target context $\mathbf{X}^Q$ is drawn form the target context distribution Q, with density $q(\cdot)$. The treatment for $\mathbf{X}^Q$ is deterministically set as $\pi(\mathbf{X}^Q)$. Rewards are drawn from $\text{Bernoulli}\big(f_{\boldsymbol{\omega}}\big(\mathbf{X}^Q, \pi(\mathbf{X}^Q)\big)\big)$, with likelihood $\vartheta_{f_{\boldsymbol{\omega}}}\big(\cdot\,; \mathbf{X}^Q, \pi(\mathbf{X}^Q)\big)$.

The following lemma offers a method to compute the KL divergence between the distributions of two problem instances.

Lemma 19 (Divergence decomposition)**.** *Fix a policy π. Fix one $\boldsymbol{\omega} \in \Omega_m$. For any $\boldsymbol{\omega}' \in \Omega$ such that $\boldsymbol{\omega}' \neq \boldsymbol{\omega}$, The* KL *divergence between $\theta_{\boldsymbol{\omega}',\pi}$ and $\theta_{\boldsymbol{\omega},\pi}$ is decomposed as follows:*

$$\mathrm{KL}(\theta_{\boldsymbol{\omega}',\pi}, \theta_{\boldsymbol{\omega},\pi}) = \mathrm{KL}_Q + \mathrm{KL}_P,$$

where

$$\mathrm{KL}_P := E_{\boldsymbol{\omega}'}\left[\sum_{t=1}^n \log\left(\frac{\vartheta_{f_{\boldsymbol{\omega}'}}(Y_t^P\,; \mathbf{X}_t^P, p_t^P)}{\vartheta_{f_{\boldsymbol{\omega}}}(Y_t^P\,; \mathbf{X}_t^P, p_t^P)}\right)\right],$$

and

$$\mathrm{KL}_Q := E_{\boldsymbol{\omega}'}\left[\log\left(\frac{\vartheta_{f_{\boldsymbol{\omega}'}}\big(Y^Q\,; \mathbf{X}^Q, \pi(\mathbf{X}^Q)\big)}{\vartheta_{f_{\boldsymbol{\omega}}}\big(Y^Q\,; \mathbf{X}^Q, \pi(\mathbf{X}^Q)\big)}\right)\right].$$

Proof of Lemma 19 is provided in Appendix D.6.11.

To use Fano's method, now we have to bound each of KL_P and KL_Q.

Lemma 20. *The KL divergences defined in Lemma 19 are bounded from above as follows:*

$$KL_P \lesssim m\,\zeta\,n\,r^{2\beta+1+\alpha+d}, \quad KL_Q \lesssim m r^{d+1+2\beta}. \tag{35}$$

The proof of Lemma 20 is provided in Appendix D.6.12.

We chose the radius parameter r as

$$r = (\zeta n)^{-1/(d+1+2\beta+\alpha)},$$

as we did in Appendix D.4.1. Then we can further bound KL_P as follows:

$$\mathrm{KL}_P \lesssim m\,\zeta\,n\,\Big((\zeta n)^{-1/(d+1+2\beta+\alpha)}\Big)^{d+1+2\beta+\alpha} = m\,\zeta\,n\,(\zeta n)^{-1} = m,$$

and bound KL_Q as follows:

$$\mathrm{KL}_Q \lesssim m\,\Big((\zeta n)^{-1/(d+1+2\beta+\alpha)}\Big)^{d+1+2\beta} = m(\zeta n)^{-(d+1+2\beta)/(d+1+2\beta+\alpha)} \lesssim m.$$

The final inequality holds because since $d, \beta \geq 0$ and $\alpha > 0$, the exponent $-(d+1+2\beta)/(d+1+2\beta+\alpha)$ is negative, and the term $(\zeta n)^{-(d+1+2\beta)/(d+1+2\beta+\alpha)}$ is therefore bounded by 1 for $\zeta n > 1$.

By Lemmas 19 and 20, we have

$$\mathrm{KL}(\theta_{\boldsymbol{\omega}',\pi}, \theta_{\boldsymbol{\omega},\pi}) \lesssim m\,\zeta\, n\, r^{d+1+2\beta+\alpha} + m r^{d+1+2\beta} \lesssim m.$$

The average KL divergence over the set of problem instances is:

$$\frac{1}{M^2} \sum_{i\in\Omega_m, j\in\Omega_m} \mathrm{KL}(\theta_{\boldsymbol{\omega},\pi}, \theta_{\boldsymbol{\omega}',\pi}) \lesssim \frac{1}{M^2} \sum_{i\in\Omega_m, j\in\Omega_m} m = m \lesssim \log_2(M)$$

where the last inequality, up to a constant, uses $\log_2 M \geq m/8$ from (20).

Based on the small KL divergence of our selected problem instances (as shown above) and their parameter separation of $(\zeta n)^{-\beta/(d+1+2\beta+\alpha)}$ (as shown in (34)), Lemma 8 gives us:

$$\inf_{\pi} \sup_{\theta_{\boldsymbol{\omega}} \in \mathcal{H}_{\Omega_m}} \mathbf{E}\left[f(\mathbf{X}^Q, t^*(\mathbf{X}^Q)) - f(\mathbf{X}^Q, \pi(\mathbf{X}^Q))\right] \gtrsim (\zeta n)^{-\frac{\beta}{2\beta+\alpha+d+1}}.$$

This concludes the proof of equation (7) of Theorem 2.

D.6 Proof of the Supporting Lemmas

D.6.1 Proof of Lemma 7

Proof. To bound the expectations, we apply the law of iterated expectations. Specifically, we first condition on the target context $\mathbf{X}^Q$ and then analyze the inner expectation with respect to the source context-treatment pairs $(\mathbf{X}_i^P, t^P)_{i=1}^n$, which are drawn i.i.d. according to μ. Conditioned on $\mathbf{X}^Q$, the optimal treatment $t^\dagger(\mathbf{X}^Q)$ becomes deterministic. In this case, the quantity $N_h\big(\mathbf{X}^Q, t^\dagger(\mathbf{X}^Q)\big)$, which is a transformation of $(\mathbf{X}_i^P, t^P)_{i=1}^n$, follows a binomial distribution with number of trials n and success probability

$$\sigma(\mathbf{X}^Q) := \mu\big(B_h(\mathbf{X}^Q) \times [t^\dagger(\mathbf{X}^Q) - h, t^\dagger(\mathbf{X}^Q) + h]\big).$$

By Assumption 4, we have the lower bound

$$\sigma(\mathbf{X}^Q) \geq 2\zeta h P\left(B_h(\mathbf{X}^Q)\right). \tag{36}$$

Using this distribution and bound, we now proceed to analyze E_1. We apply Lemma 4 to bound the ratio of an indicator function of a binomial random variable to the value of that random variable:

$$\begin{aligned}
E_1 &= \mathbb{E}_Q\left[\mathbb{E}_\mu\left[\frac{\mathbb{1}\{N_h\big(\mathbf{X}^Q, t^\dagger(\mathbf{X}^Q)\big) > 0\}}{\sqrt{N_h\big(\mathbf{X}^Q, t^\dagger(\mathbf{X}^Q)\big)}} \mid \mathbf{X}^{\mathbf{Q}}\right]\right] \\
&\overset{(i)}{=} \mathbb{E}_Q\left[\mathbb{E}_\mu\left[\sqrt{\frac{\mathbb{1}\{N_h\big(\mathbf{X}^Q, t^\dagger(\mathbf{X}^Q)\big) > 0\}}{N_h\big(\mathbf{X}^Q, t^\dagger(\mathbf{X}^Q)\big)}} \mid \mathbf{X}^{\mathbf{Q}}\right]\right] \\
&\overset{(ii)}{\leq} \sqrt{\mathbb{E}_Q\left[\mathbb{E}_\mu\left[\frac{\mathbb{1}\{N_h\big(\mathbf{X}^Q, t^\dagger(\mathbf{X}^Q)\big) > 0\}}{N_h\big(\mathbf{X}^Q, t^\dagger(\mathbf{X}^Q)\big)} \mid \mathbf{X}^{\mathbf{Q}}\right]\right]} \\
&\overset{(iii)}{\lesssim} \sqrt{\mathbb{E}_Q\left[\frac{1}{n\sigma(\mathbf{X}^Q)}\right]} \\
&\overset{(iv)}{\leq} \sqrt{\mathbb{E}_Q\left[\frac{1}{2n\zeta h P\left(B_h(\mathbf{X}^Q)\right)}\right]} \\
&\overset{(iv)}{=} \frac{1}{\sqrt{2n\zeta h}} \sqrt{\mathbb{E}_Q\left[\frac{1}{P\left(B_h(\mathbf{X}^Q)\right)}\right]}.
\end{aligned} \tag{37}$$

where each step uses:

(i) the indicator is the same as its square root,

(ii) repeated application of the Jensen's inequality,

(iii) Lemma 4,

(iv) the bound in (36).

All that remains is to control the expectation

$$\mathbb{E}_Q\left[\frac{1}{P\left(B_h(\mathbf{X}^Q)\right)}\right].$$

To do this, we invoke the transfer exponent assumption (Assumption 3).Let $\mathcal{N}$ denote a a minimal $h/2$-net of $[0,1]^d$. Using the fact that $|\mathcal{N}| \lesssim h^{-d}$, we have

$$\begin{aligned}
\mathbb{E}_Q\left[\frac{1}{P(B_h(\mathbf{X}^Q))}\right] &= \int_{[0,1]^d} \frac{1}{P\big(B_h(\mathbf{x})\big)}\, dQ(\mathbf{x}) \\
&\overset{(i)}{\lesssim} \int_{[0,1]^d} \frac{h^{-\alpha}}{Q\big(B_h(\mathbf{x})\big)}\, dQ(\mathbf{x}) \\
&\overset{(ii)}{\le} h^{-\alpha} \int_{\mathcal{N}} \frac{1}{Q\big(B_h(\mathbf{x})\big)}\, dQ(\mathbf{x}) \\
&= h^{-\alpha} \sum_{\mathbf{z}\in\mathcal{N}} \int_{B_{h/2}(\mathbf{z})} \frac{1}{Q(B_h(\mathbf{x}))}\, dQ(\mathbf{x}) \\
&\overset{(iii)}{\le} h^{-\alpha} \sum_{\mathbf{z}\in\mathcal{N}} \int_{B_{h/2}(\mathbf{z})} \frac{1}{Q\big(B_{h/2}(\mathbf{z})\big)}\, dQ(\mathbf{x}) \\
&= h^{-\alpha} \sum_{\mathbf{z}\in\mathcal{N}} \frac{1}{Q\big(B_{h/2}(\mathbf{z})\big)} \int_{B_{h/2}(\mathbf{z})} dQ(\mathbf{x}) \\
&= h^{-\alpha} \sum_{\mathbf{z}\in\mathcal{N}} \frac{1}{Q\big(B_{h/2}(\mathbf{z})\big)} Q\big(B_{h/2}(\mathbf{z})\big) \\
&= h^{-\alpha} \cdot |\mathcal{N}| \\
&\lesssim h^{-(d+\alpha)},
\end{aligned} \tag{38}$$

where each step uses:

(i) $\kappa(P,Q) \le \alpha$,

(ii) $[0,1]^d \subset \bigcup_{\mathbf{z}\in\mathcal{N}} B_{h/2}(\mathbf{z})$

(iii) $\mathbf{x} \in B_{h/2}(\mathbf{z})$ implies $B_{h/2}(\mathbf{z}) \subset B_h(\mathbf{x})$, thus we have $1/Q\big(B_h(\mathbf{x})\big) \le 1/Q\big(B_{h/2}(\mathbf{z})\big)$

Therefore, combining (37) and (38), we have the following bound for E_1:

$$E_1 \lesssim (\zeta\, n\, h)^{-1/2} h^{-(d+\alpha)/2}. \tag{39}$$

□

D.6.2 Proof of Lemma 8

Proof. We have for any policy π,

$$\begin{aligned}
\sup_{\theta\in\Theta} E_{D\sim P_\theta}[L(\theta,\pi(D))] &\ge \sup_{i\in[m]} E_{D\sim P_{\theta_i}}[L(\theta_i,\pi(D))] \\
&\ge \frac{1}{m}\sum_{i=1}^{m} E_{D\sim P_{\theta_i}}[L(\theta_i,\pi(D))].
\end{aligned}$$

Now denoting $\Psi(D) \equiv \arg\min_{i\in[m]} L(\theta_i, \pi(D))$, we get

$$L(\theta_i, \pi(D)) \geq \frac{1}{2}(L(\theta_i, \pi(D)) + L(\theta_{\Psi(D)}, \pi(D))) \geq \frac{\Delta}{2}\mathbb{I}\{\Psi(D) \neq i\},$$

where $\mathbb{I}$ is the indicator function. Thus,

$$E_{D\sim P_{\theta_i}}[L(\theta_i, \pi(D))] \geq \frac{\Delta}{2} P_{\theta_i}(\Psi \neq i),$$

taking summation over i, we get the first inequality. The second inequality follows from Proposition 15.12 and equation (15.34) of Wainwright [27]. □

D.6.3 Proof of Lemma 9

Proof. The proof follows the same steps as the proof in Appendix C.1 of Wang et al. [28]. We include it here to show that the transfer component assumption holds even with our different definition of $q(\mathbf{x})$ in (22), which does not incorporate the exploration condition.

Assumption 3 requires that its condition holds for all $\mathbf{x}$ in the support of Q. Given that the equation (22) constructs the support of Q to be the union of the set $\bigcup_{i=1}^m B_{r/4}(\mathbf{x}_i^*)$ and $\mathcal{X} \setminus \bigcup_{i=1}^m B_r(\mathbf{x}_i^*)$, we only need to verify the assumption for $\mathbf{x}$ in these two non-overlapping regions. First we consider a small radius of $h \leq 3r/4$ for each of these two regions:

The case where $h \leq 3r/4$ and $\mathbf{x}$ is contained in a small ball. In this scenario, $\mathbf{x} \in B_{r/4}(\mathbf{x}_i^*)$ for some $i \in [m]$. By the construction in (22), the density $q(\mathbf{x})$ is zero in the annulus between the $r/4$-ball and the r-ball. This means the measure of the ball $B_h(\mathbf{x})$ under Q is given by:

$$Q(B_h(\mathbf{x})) = q_1 \text{Leb}\left(B_h(\mathbf{x}) \cap B_{r/4}(\mathbf{x}_i^*)\right). \tag{40}$$

On the other hand, the construction in (23) shows that the density $p(\mathbf{x})$ s non-negative (specifically, δ) in the annulus between the $r/2$-ball and the r-ball. It follows that the measure of $B_h(\mathbf{x})$ under P is bounded as follows:

$$P\big(B_h(\mathbf{x})\big) \geq C_\alpha r^\alpha q_1 \text{Leb}\left(B_h(\mathbf{x}) \cap B_{r/4}(\mathbf{x}_i^*)\right) \overset{(i)}{\geq} C_\alpha r^\alpha Q(B_h(\mathbf{x})) \overset{(ii)}{\geq} C_\alpha h^\alpha Q(B_h(\mathbf{x})),$$

where step (i) uses (40) and step (ii) uses $h \leq 3r/4$.

The case of $h \leq 3r/4$ and $\mathbf{x}$ is not contained any large ball. In this scenario, we have $\mathbf{x} \in \mathcal{X} \setminus \bigcup_{j=1}^m B_r(\mathbf{x}_i^*)$. This means that $B_h(\mathbf{x})$ does not overlap with any of $B_{r/4}(\mathbf{x}_j^*)$'s. By the construction in (22), this means the measure of the ball $B_h(\mathbf{x})$ under Q is given by:

$$Q\big(B_h(\mathbf{x})\big) = q_0 \text{ Leb}\bigg(B_h(\mathbf{x}) \cap \big(\mathcal{X} \setminus \bigcup_{j=1}^m B_r(\mathbf{x}_j^*)\big)\bigg). \tag{41}$$

On the other hand, the construction in (23) shows that the density $p(\mathbf{x})$ s non-negative (specifically, δ) in the annulus between the $r/2$-ball and the r-ball. It follows that we have

$$P\big(B_h(\mathbf{x})\big) \geq Q\big(B_h(\mathbf{x})\big) \geq C_\alpha h^\alpha Q\big(B_h(\mathbf{x})\big),$$

where the last inequality uses $h \leq 3r/4$ and $0 < C_\alpha \leq 1$.

The case of $3r/4 < h \leq 1$. Since the constructions of $q(\mathbf{x})$ and $p(\mathbf{x})$ in (22) and (23) implies that $P\big(B_r(\mathbf{x}_j^*)\big) = Q\big(B_r(\mathbf{x}_j^*)\big)$ for all $j \in [m]$. We can leverage this to verify that the above inequalities hold for $3r/4 < h \leq 1$.

Combining the results from all cases, the transfer exponent of the constructed source context distribution with respect to the target context distribution is α, with the corresponding constant C_α.

□

D.6.4 Proof of Lemma 9

Proof. The proof follows the same steps as the proof in Appendix C.1 of Wang et al. [28]. We include it here to show that the transfer component assumption holds even with our weaker definition of exploration condition in Assumption 4.

Pick $t^\dagger(\mathbf{x})$ as the minimum optimal treatment defined in (30). As shown in Lemma 14, this minimum value is guaranteed to exist. Recall that our construction of $\wp(p \mid \mathbf{x})$ in equation (24) does not depend on $\mathbf{x}$. Therefore, for any $\mathbf{x} \in \mathcal{X}$ and a given $r \in (0, 1]$, we have that:

$$\mu([t^\dagger(\mathbf{x}) - r, t^\dagger(\mathbf{x}) + r] \times B_r(\mathbf{x})) = P\big(B_r(\mathbf{x})\big)\, 2r \, \min\left\{\zeta, \frac{1 - r\zeta}{1 - r}\right\} = P\big(B_r(\mathbf{x})\big)\, 2r\zeta,$$

where the last equality holds because

$$\zeta - \frac{1 - r\zeta}{1 - r} = \frac{\zeta(1 - r) - (1 - r\zeta)}{1 - r} = \frac{\zeta - r\zeta - 1 + r\zeta}{1 - r} = \frac{\zeta - 1}{1 - r} \leq 0.$$

As a result, the exploration coefficient defined in Assumption 4 for the constructed source context-treatment pair distribution equals ζ. □

D.6.5 Proof of Lemma 11

Proof. To prove that the function $\phi(z)$ is Hölder continuous, we need to show that there exists a constant C such that for any $z_1, z_2 \in \mathbb{R}_+$, we have $|\phi(z_1) - \phi(z_2)| \leq C|z_1 - z_2|^\beta$. We will analyze this by considering the different intervals where z_1 and z_2 can be.

Case 1: $z_1, z_2 \in [1/4, 1/2]$. In this interval, the function is defined as $\phi(z) = (2 - 4z)^\beta$. We will use the inequality $|a^\beta - b^\beta| \leq |a - b|^\beta$ for $a, b \geq 0$ and $\beta \in (0, 1]$. Let $a = 2 - 4z_1$ and $b = 2 - 4z_2$. Since $z_1, z_2 \in [1/4, 1/2]$, we have $a, b \in [0, 1]$, so the inequality applies.

$$\begin{aligned}
|\phi(z_1) - \phi(z_2)| &= |(2 - 4z_1)^\beta - (2 - 4z_2)^\beta| \\
&\leq |(2 - 4z_1) - (2 - 4z_2)|^\beta \\
&= |4z_2 - 4z_1|^\beta \\
&= |4(z_2 - z_1)|^\beta \\
&= 4^\beta |z_1 - z_2|^\beta.
\end{aligned}$$

This shows that on the interval $[1/4, 1/2]$, the function is Hölder continuous with constant 4^β.

Case 2: z_1, z_2 are in different intervals. We now consider the cases where z_1 and z_2 are not in the same interval. Without loss of generality, assume $z_1 < z_2$. The only nontrivial cases are when one point is in an interval where the function is constant and the other is not.

Subcase 2.1: $z_1 < 1/4$ and $z_2 \in [1/4, 1/2]$. Here, $\phi(z_1) = 1$ and $\phi(z_2) = (2 - 4z_2)^\beta$. We have

$$\begin{aligned}
|\phi(z_1) - \phi(z_2)| &= 1 - (2 - 4z_2)^\beta \\
&\leq 1 - (2 - 4(z_1 + d))^\beta \\
&= 1 - (2 - 4z_1 - 4d)^\beta
\end{aligned}$$

where $d = z_2 - z_1 > 0$. A simpler approach is to use the fact that the maximum value of the function's derivative is at the endpoints of the intervals. We have $|\phi(z_1) - \phi(z_2)| = 1 - (2 - 4z_2)^\beta$ and $|z_1 - z_2| = z_2 - z_1$. Since $z_1 < 1/4 \leq z_2$, we have $z_2 - z_1 > z_2 - 1/4 = (4z_2 - 1)/4$. This implies $4(z_2 - z_1) > 4z_2 - 1$. The function $f(x) = 1 - x^\beta$ on $[0, 1]$ is convex, so $1 - x^\beta \leq \beta(1 - x)$ for $\beta \leq 1$. Let's use the inequality $1 - x^\beta \leq (1 - x)^\beta$ for $x \in [0, 1]$. Let $x = 2 - 4z_2 \in [0, 1]$.

$$|\phi(z_1) - \phi(z_2)| = 1 - (2 - 4z_2)^\beta \leq (1 - (2 - 4z_2))^\beta = (4z_2 - 1)^\beta$$

Since $z_2 - z_1 \geq z_2 - 1/4$, we have $4(z_2 - z_1) \geq 4z_2 - 1$, so $4^\beta |z_2 - z_1|^\beta \geq (4z_2 - 1)^\beta$. Thus, $|\phi(z_1) - \phi(z_2)| \leq 4^\beta |z_1 - z_2|^\beta$.

Subcase 2.2: $z_1 \in [1/4, 1/2]$ **and** $z_2 > 1/2$**.** Here, $\phi(z_1) = (2 - 4z_1)^\beta$ and $\phi(z_2) = 0$. We have $|\phi(z_1) - \phi(z_2)| = (2 - 4z_1)^\beta$. Since $z_1 \in [1/4, 1/2]$ and $z_2 > 1/2$, we have $|z_1 - z_2| = z_2 - z_1$. Also, $z_2 - z_1 > 1/2 - z_1 = (2 - 4z_1)/4$. This implies $4(z_2 - z_1) > 2 - 4z_1$. Taking the β-th power of both sides, we get $4^\beta (z_2 - z_1)^\beta > (2 - 4z_1)^\beta = |\phi(z_1) - \phi(z_2)|$. Thus, $|\phi(z_1) - \phi(z_2)| \leq 4^\beta |z_1 - z_2|^\beta$.

Subcase 2.3: $z_1 < 1/4$ **and** $z_2 > 1/2$**.** In this case, $|\phi(z_1) - \phi(z_2)| = |1 - 0| = 1$. Also, since $z_1 < 1/4$ and $z_2 > 1/2$, we have $|z_1 - z_2| = z_2 - z_1 > 1/2 - 1/4 = 1/4$. Taking the β-th power, $|z_1 - z_2|^\beta > (1/4)^\beta$. Multiplying by 4^β, we get $4^\beta |z_1 - z_2|^\beta > 4^\beta (1/4)^\beta = 1$. So, $|\phi(z_1) - \phi(z_2)| = 1 < 4^\beta |z_1 - z_2|^\beta$.

In all cases, the inequality $|\phi(z_1) - \phi(z_2)| \leq 4^\beta |z_1 - z_2|^\beta$ holds. Therefore, $\phi(z)$ is Hölder continuous with exponent β and Hölder constant 4^β. □

D.6.6 Proof of Lemma 12

Proof. By the construction in (29), it suffices to show that the function $\varphi_\beta : \mathcal{X} \times \mathcal{T} \to [0, 1/4]$ satisfies Assumption 1 with respect to the Hölder constant $C_\mathrm{H} > 0$. For any $(\mathbf{x}, t), (\mathbf{x}', t') \in \mathcal{X} \times \mathcal{T}$, we have that

$$|\varphi_\beta(\mathbf{x}, t) - \varphi_\beta(\mathbf{x}', t')| = C_{\varphi_\beta} r \big|\phi\big(\max\{\|\mathbf{x}\|_\infty, |p|\}/r\big) - \phi\big(\max\{\|\mathbf{x}'\|_\infty, |t'|\}/r\big)\big| \tag{42}$$

Since ϕ is a Hölder continuous function with Hölder constant 4^β, it follows that

$$\begin{aligned}
|\varphi_\beta(\mathbf{x}, t) - \varphi_\beta(\mathbf{x}', t')| &\leq 4^\beta C_\psi r \big|\max\{\|\mathbf{x}\|_\infty, |p|\}/r - \max\{\|\mathbf{x}'\|_\infty, |t'|\}/r\big| \\
&= 4^\beta C_{\varphi_\beta} \left|\max\{\|\mathbf{x}\|_\infty, |p|\} - \max\{\|\mathbf{x}'\|_\infty, |t'|\}\right| \\
&\overset{(i)}{\leq} 4^\beta C_{\varphi_\beta} \max\big\{\big|\|\mathbf{x}\|_\infty - \|\mathbf{x}'\|_\infty\big|, \big||p| - |t'|\big|\big\} \\
&\overset{(ii)}{\leq} 4^\beta C_{\varphi_\beta} \max\big\{\|\mathbf{x} - \mathbf{x}'\|_\infty, |p - t'|\big\} \\
&\leq C_\mathrm{H} \max\big\{\|\mathbf{x} - \mathbf{x}'\|_\infty, |p - t'|\big\},
\end{aligned}$$

where step (i) uses a basic inequality $|\max(a, b) - \max(c, d)| \leq \max(|a - c|, |b - d|)$, step (ii) uses the reverse triangle inequality, and step (iii) holds since $C_{\varphi_\beta} = (C_\mathrm{H}/4^\beta) \wedge (1/4^\beta)$. Thus, for any $\boldsymbol{\omega} \in \Omega_m$, the reward function $f_{\boldsymbol{\omega}}$ satisfies Assumption 1 with respect to the Hölder constant $C_\mathrm{H} > 0$. □

D.6.7 Proof of Lemma 14

Proof. For any $\boldsymbol{\omega} \in \Omega_m$, by our construction, the reward function $f_{\boldsymbol{\omega}}(\mathbf{x}, t)$ is continuous with respect to p for any fixed $\mathbf{x} \in \mathcal{X}$. Since the domain for the treatment, $\mathcal{T} = [0, 1]$, is a compact set, the Extreme Value Theorem guarantees that a maximum value, M, is attained. The set of all treatments that achieve this maximum value, known as the $\arg\max$ set, can be expressed as the preimage of the closed set $\{M\}$ under the continuous function $f_{\boldsymbol{\omega}}(\mathbf{x}, \cdot)$, i.e., $\{t \in \mathcal{T} \mid f_{\boldsymbol{\omega}}(\mathbf{x}, t) = M\}$. Since the preimage of a closed set by a continuous function is still closed, $\arg\max$ set is a closed subset of the compact domain $\mathcal{T}$, and thus is also compact. Finally, because the $\arg\max$ set is a non-empty compact set of real numbers, it is guaranteed to contain a minimum element. Therefore, the minimal optimal treatment, $p^\dagger_{\boldsymbol{\omega}}(\mathbf{x})$, is well-defined. □

D.6.8 Proof of Lemma 15

Proof. Since $\mathbf{x} \in B_{[0,1]^d}(\mathbf{x}_i^*, r/4)$ for some $i \in [m]$, by definition of the L_∞-norm ball, we have $\|\mathbf{x} - \mathbf{x}_i^*\|_\infty < r/4$. Due to the $2r$-spacing of the grid points $S_{[0,1]^d, r}$ and the localization of the reward function's bumps within r-balls, the reward function simplifies to a single bump. We now analyze the two cases for ω_i:

Case 1: $\omega_i = -1$**.** The reward function becomes $f_{\boldsymbol{\omega}}(\mathbf{x}, t) = \frac{1}{2} - \varphi_\beta\left(\mathbf{x} - \mathbf{x}_i^*, p - \tilde{p}\right)$. To maximize this function, we must minimize the non-negative term $\varphi_\beta\left(\mathbf{x} - \mathbf{x}_i^*, p - \tilde{p}\right)$. The function φ_β, defined in (26) is zero when its argument $z := \frac{1}{r}\max\{\|\mathbf{x} - \mathbf{x}_i^*\|_\infty, |p - \tilde{p}|\}$ satisfies $z \geq 1/2$. This is equivalent to $\max\{\|\mathbf{x} - \mathbf{x}_i^*\|_\infty, |p - \tilde{p}|\} \geq r/2$. Given that $\|\mathbf{x} - \mathbf{x}_i^*\|_\infty < r/4$, this condition simplifies to requiring $|p - \tilde{p}| \geq r/2$. We need to find the lowest treatment p that satisfies this condition. As $\tilde{p}$ belongs to the grid point set $S_{\mathcal{T},r}$, which starts from $r/2$, it holds that $\tilde{p} \geq r/2$. Consequently $p = 0$ is the smallest $t \in \mathcal{T}$ that satisfies $|p - \tilde{p}| \geq r/2$, since $|0 - \tilde{p}| = \tilde{p} \geq r/2$. Therefore 0 is the optimal treatment.

Case 2: $\omega_i = 1$**.** The reward function becomes $f_{\boldsymbol{\omega}}(\mathbf{x}, t) = \frac{1}{2} + \varphi_\beta\left(\mathbf{x} - \mathbf{x}_i^*,\, p - \tilde{p}\right)$. To maximize $f_{\boldsymbol{\omega}}(\mathbf{x}, t)$, we must maximize $\varphi_\beta\left(\mathbf{x} - \mathbf{x}_i^*,\, p - \tilde{p}\right)$. The maximum value of φ_β is achieved when its argument $z = \frac{1}{r} \max\{\|\mathbf{x} - \mathbf{x}_i^*\|_\infty, |p - \tilde{p}|\}$ satisfies $0 \leq z \leq 1/4$. So, we need $\max\{\|\mathbf{x} - \mathbf{x}_i^*\|_\infty, |p - \tilde{p}|\} \leq r/4$. Since $\|\mathbf{x} - \mathbf{x}_i^*\|_\infty < r/4$, the condition simplifies to $|p - \tilde{p}| \leq r/4$. The smallest such p is $\tilde{p} - r/4$, if it is nonnegative. Since $\tilde{p} \geq r/2$, we have $\tilde{p} - r/4 \geq r/4 \geq 0$, ensuring this is a valid non-negative treatment. Therefore $\tilde{p} - r/4$ is the optimal treatment. This completes the proof of Lemma 15. □

D.6.9 Proof of Lemma 17

Proof. We first note that $\mathbf{x} \in B_{[0,1]^d}(\mathbf{x}_i^*, r/4)$ for some $i \in [m]$. If $\omega_i = -1$, we have $p^\dagger_{\boldsymbol{\omega}}(\mathbf{x}) = 0$ by Lemma 15. Since $0 \notin (\tilde{p} - r/2, \tilde{p} + r/2]$, we have $h(p^\dagger_{\boldsymbol{\omega}}(\mathbf{x})) = 0$. Therefore,

$$\begin{aligned} f_{\boldsymbol{\omega}}(\mathbf{x}, p^\dagger_{\boldsymbol{\omega}}(\mathbf{x})) - f_{\boldsymbol{\omega}}(\mathbf{x}, t) &\overset{(i)}{=} C_\phi r^\beta \mathbb{1}\{|p - \tilde{p}| \leq r/2\} \\ &= C_\phi r^\beta h(p) \\ &= C_\phi r^\beta \mathbb{1}\{h(p^\dagger_{\boldsymbol{\omega}}(\mathbf{x})) \neq h(p)\}. \end{aligned}$$

where step (i) uses $p \in \bigcup_{j=1}^{\lfloor 1/r \rfloor} [p_j^* - r/4, p_j^* + r/4]$.

Next, if $\omega_i = 1$, we have $p^\dagger_{\boldsymbol{\omega}}(\mathbf{x}) = \tilde{p} - r/4$ by Lemma 15, and $h(p^\dagger_{\boldsymbol{\omega}}(\mathbf{x})) = 1$. Hence,

$$\begin{aligned} f_{\boldsymbol{\omega}}(\mathbf{x}, p^\dagger_{\boldsymbol{\omega}}(\mathbf{x})) - f_{\boldsymbol{\omega}}(\mathbf{x}, t) &= C_\phi r^\beta [1 - \mathbb{1}\{|p - \tilde{p}| \leq r/2\}] \\ &= C_\phi r^\beta [1 - h(p)] \\ &= C_\phi r^\beta \mathbb{1}\{h(p^\dagger_{\boldsymbol{\omega}}(\mathbf{x})) \neq h(p)\}. \end{aligned}$$

□

D.6.10 Proof of Lemma 18

Proof. Let us assume that $\pi(\mathbf{x}) \in [p_i^* - r/2, p_i^* + r/2] \setminus [p_i^* - r/4, p_i^* + r/4]$ for some $i \in [\lfloor 1/r \rfloor]$. We will analyze this situation by considering two primary cases based on the location of $\mathbf{x}$.

Case 1: $\mathbf{x} \in [0,1]^d \setminus \bigcup_{l=1}^m B_X(x_l^*, r/2)$ If $\mathbf{x}$ lies outside the union of the balls $B_X(x_i^*, r/2)$, then, in accordance with Lemma 13, the reward function $f_{\boldsymbol{\omega}}(\mathbf{x}, t)$ evaluates to $1/2$ for all $\boldsymbol{\omega} \in \Omega_m$ and for any treatment $p \in [0, 1]$. This implies that the specific treatment chosen has no influence on the resulting reward. Therefore, we are at liberty to select $\pi'(\mathbf{x})$ such that it falls within the desired interval:

$$\pi'(\mathbf{x}) \in [p_i^* - r/4, p_i^* + r/4].$$

Case 2: $\mathbf{x} \in \bigcup_{l=1}^m B_X(x_l^*, r/2)$ We consider the cases of $\omega_i = 1$ and $\omega_i = -1$.

Case 2.1: $\omega_i = 1$ We proceed by examining two distinct sub-scenarios:

- **If** $p_i^* \neq \tilde{p}$**:** Recall that the treatment grids $\mathcal{S}_{\mathcal{T},r}$, defined in (17), have r-spacing. Thus for any treatment $p \in [p_i^* - r/2, p_i^* + r/2]$, we have $|\tilde{p} - p| \geq r/2$. Thus by Lemma 13, the reward function value remains the same for any $p \in [p_i^* - r/2, p_i^* + r/2]$. Consequently, by setting $\pi'(\mathbf{x}) \in [p_i^* - r/4, p_i^* + r/4]$, the reward associated with π' will be identical to that of π.
- **If** $p_i^* = \tilde{p}$**:** The positive bump of the reward function, formally represented by φ_β as defined in (26) and (29), is a non-increasing function of the distance between p and t_i^*. Thus, if we choose
$$\pi'(\mathbf{x}) \in [p_i^* - r/4, p_i^* + r/4],$$
this selection positions $\pi'(\mathbf{x})$ closer to $\tilde{p}$ than $\pi(\mathbf{x})$. Consequently, the reward function value for π' will be no greater than that for π.

Case 2.2: $\omega_i = -1$ We proceed by examining two distinct sub-scenarios:

- **If** $p_i^* \neq \tilde{p}$**:** Consistent with the reasoning in Case 2.1, we can safely choose $\pi'(\mathbf{x}) \in [p_i^* - r/4, p_i^* + r/4]$.
- **If** $p_i^* = \tilde{p}$**:** The negative bump of the reward function, formally represented by $-\varphi_\beta$ as defined in (26) and (29), is a non-decreasing function of the distance between p and p_i^*. Thus, if we opt for

  $$\pi'(\mathbf{x}) \in [p_j^* - r/4, p_j^* + r/4], \quad \text{for some } j \neq i,$$

  this selection places $\pi'(\mathbf{x})$ further from $\tilde{p}$ than $\pi(\mathbf{x})$. As a result, the reward function value for π' will be no greater than that for π. This completes the proof of Lemma 18.

□

D.6.11 Proof of Lemma 19

Proof. Let $\vartheta_{\boldsymbol{\omega},\pi}$ the density corresponding to the distribution $\vartheta_{\boldsymbol{\omega},\pi}$ defined in Definition 2. It can be expanded as

$$\vartheta_{\boldsymbol{\omega},\pi}(\mathcal{D}^P, \mathcal{D}^Q) = \underbrace{q(\mathbf{X}^Q)\, \vartheta_{f_{\boldsymbol{\omega}}}\left(Y^Q\,;\, \mathbf{X}^Q, \pi(\mathbf{X}^Q)\right)}_{\text{Corresponds to } \mathcal{D}^Q} \underbrace{\left\{\prod_{t=1}^{n} p(\mathbf{X}_t^P)\, \wp(p_t^P|\mathbf{X}_t^P)\, \vartheta_{f_{\boldsymbol{\omega}}}(Y_t^P\,;\, \mathbf{X}_t^P, p_t^P)\right\}}_{\text{Corresponds to } \mathcal{D}^P}$$

For any $\boldsymbol{\omega}' \in \Omega_m$ and a reference $\boldsymbol{\omega} \in \Omega_m$, we have the following computation, due to cancellations:

$$\log\left(\frac{d\theta_{\boldsymbol{\omega}',\pi}}{d\theta_{\boldsymbol{\omega},\pi}}(\mathcal{D}^P, \mathcal{D}^Q)\right) = \log\left(\frac{\vartheta_{f_{\boldsymbol{\omega}'}}\left(Y^Q\,;\, \mathbf{X}^Q, \pi(\mathbf{X}^Q)\right)}{\vartheta_{f_{\boldsymbol{\omega}}}\left(Y^Q\,;\, \mathbf{X}^Q, \pi(\mathbf{X}^Q)\right)}\right) + \sum_{t=1}^{n} \log\left(\frac{\vartheta_{f_{\boldsymbol{\omega}'}}(Y_t^P\,;\, \mathbf{X}_t^P, p_t^P)}{\vartheta_{f_{\boldsymbol{\omega}}}(Y_t^P\,;\, \mathbf{X}_t^P, p_t^P)}\right).$$

Taking the expectation with respect to $\theta_{\boldsymbol{\omega}',\pi}$, we obtain

$$\begin{aligned}
\mathrm{KL}(\theta_{\boldsymbol{\omega}',\pi}, \theta_{\boldsymbol{\omega},\pi}) &= \mathbb{E}_{\boldsymbol{\omega}'}\left[\log\left(\frac{d\theta_{\boldsymbol{\omega}',\pi}}{d\theta_{\boldsymbol{\omega},\pi}}(\mathcal{D}^P, \mathcal{D}^Q)\right)\right] \\
&= \mathbb{E}_{\boldsymbol{\omega}'}\left[\log\left(\frac{\vartheta_{f_{\boldsymbol{\omega}'}}\left(Y^Q\,;\, \mathbf{X}^Q, \pi(\mathbf{X}^Q)\right)}{\vartheta_{f_{\boldsymbol{\omega}}}\left(Y^Q\,;\, \mathbf{X}^Q, \pi(\mathbf{X}^Q)\right)}\right)\right] + \mathbb{E}_{\boldsymbol{\omega}'}\left[\sum_{t=1}^{n} \log\left(\frac{\vartheta_{f_{\boldsymbol{\omega}'}}(Y_t^P\,;\, \mathbf{X}_t^P, p_t^P)}{\vartheta_{f_{\boldsymbol{\omega}}}(Y_t^P\,;\, \mathbf{X}_t^P, p_t^P)}\right)\right] \\
&= \mathrm{KL}_Q + \mathrm{KL}_P.
\end{aligned}$$

This completes the proof of Lemma 19.

□

D.6.12 Proof of Lemma 20

Proof. We start by bounding KL_P from above:

$$\begin{aligned}
\mathrm{KL}_P &= \mathbb{E}_{\boldsymbol{\omega}'}\left[\sum_{t=1}^{n} \log\left(\frac{\vartheta_{f_{\boldsymbol{\omega}'}}(Y_t^P\,;\,\mathbf{X}_t^P, p_t^P)}{\vartheta_{f_{\boldsymbol{\omega}}}(Y_t^P\,;\,\mathbf{X}_t^P, p_t^P)}\right)\right] \\
&= \sum_{t=1}^{n} \mathbb{E}_{\boldsymbol{\omega}'}\left[\log\left(\frac{\vartheta_{f_{\boldsymbol{\omega}'}}(Y_t^P\,;\,\mathbf{X}_t^P, p_t^P)}{\vartheta_{f_{\boldsymbol{\omega}}}(Y_t^P\,;\,\mathbf{X}_t^P, p_t^P)}\right)\right] \\
&= \sum_{t=1}^{n} \int_{[0,1]^d} \mathbb{E}_{\boldsymbol{\omega}'}\left[\log\left(\frac{\vartheta_{f_{\boldsymbol{\omega}'}}(Y_t^P\,;\,\mathbf{x}, p_t^P)}{\vartheta_{f_{\boldsymbol{\omega}}}(Y_t^P\,;\,\mathbf{x}, p_t^P)}\right) \big| \mathbf{X}_t^P = \mathbf{x}\right] p(\mathbf{x}) d\mathbf{x} \\
&\overset{(i)}{=} \sum_{t=1}^{n} \int_{\bigcup_{j=1}^{m} B_{r/4}(\mathbf{x}_j^*)} \mathbb{E}_{\boldsymbol{\omega}'}\left[\mathbb{1}\{p_t^P \in [\tilde{p} - r/2, \tilde{p} + r/2]\} \log\left(\frac{\vartheta_{f_{\boldsymbol{\omega}'}}(Y_t\,;\,\mathbf{x}, p_t^P)}{\vartheta_{f_{\boldsymbol{\omega}}}(Y_t\,;\,\mathbf{x}, p_t^P)}\right) \big| \mathbf{X}_t^P = \mathbf{x}\right] p(\mathbf{x}) d\mathbf{x} \\
&\overset{(ii)}{\leq} \sum_{t=1}^{n} \sum_{j \in [m] : \omega_j' \neq \omega_j} P\big(B_{r/4}(\mathbf{x}_j^*)\big)\, 2r\zeta\, \mathrm{KL}\big(\text{Bernoulli}(1/2 + C_\beta r^\beta) \,\|\, \text{Bernoulli}(1/2 - C_\beta r^\beta)\big) \\
&\overset{(iii)}{\leq} 32\, n\, r\zeta (C_\beta r^\beta)^2 \sum_{j \in [m] : \omega_j' \neq \omega_j} P\big(B_{r/4}(\mathbf{x}_j^*)\big) \\
&\overset{(iv)}{\leq} 32\, C_\beta^2\, C_\alpha\, n\, \zeta\, m\, r^{2\beta + 1 + \alpha + d} \\
&\overset{(v)}{\lesssim} \zeta m n r^{2\beta + 1 + \alpha + d},
\end{aligned}$$

where

- step (i) uses the cancellation inside the log due to Lemma 13, and the construction of $p(\mathbf{x})$ given in (23), which is zero on the annulus between the $r/2$-ball and the $r/4$-ball,
- step (ii) uses the construction of the conditional distribution $\wp(p \,|\, \mathbf{x})$, which is independent of $\mathbf{x}$, given in (24), where $\wp(p \,|\, \mathbf{x}) = \zeta$ for $p \in [\tilde{p} - r/2, \tilde{p} + r/2]$, and the construction of the reward function given in (29),
- step (iii) uses Lemma 5,
- step (iv) uses the construction of $p(\mathbf{x})$ given in (23), which is equal to $C_\alpha r^\alpha q_1 = C_\alpha r^{\alpha + d} / \mathrm{Leb}\big(B_{r/4}(\mathbf{x}_j^*)\big)$ inside the $r/4$-balls,
- step (v) uses the fact that C_α and C_β do not depend on n, m, r or ζ.

Next, we switch our gears to bounding KL_Q from above, using similar arguments as above:

$$\begin{aligned}
\mathrm{KL}_Q &= \mathbb{E}_{\boldsymbol{\omega}'}\left[\log\left(\frac{\vartheta_{f_{\boldsymbol{\omega}'}}\big(Y^Q\,;\,\mathbf{X}^Q,\pi(\mathbf{X}^Q)\big)}{\vartheta_{f_{\boldsymbol{\omega}}}\big(Y^Q\,;\,\mathbf{X}^Q,\pi(\mathbf{X}^Q)\big)}\right)\right] \\
&= \int_{[0,1]^d} \mathbb{E}_{\boldsymbol{\omega}'}\left[\log\left(\frac{\vartheta_{f_{\boldsymbol{\omega}'}}(Y^Q\,;\,\mathbf{x},\pi(\mathbf{x}))}{\vartheta_{f_{\boldsymbol{\omega}}}(Y^Q\,;\,\mathbf{x},\pi(\mathbf{x}))}\right)\big|\mathbf{X}^Q=\mathbf{x}\right] q(\mathbf{x})\,d\mathbf{x} \\
&\overset{(i)}{=} \int_{\bigcup_{j=1}^m B_{r/4}(\mathbf{x}_j^*)} E_{\boldsymbol{\omega}'}\left[\mathbb{1}\{\pi(\mathbf{x})\in[\tilde{p}-r/2,\tilde{p}+r/2]\}\log\left(\frac{\vartheta_{f_{\boldsymbol{\omega}'}}(Y^Q\,;\,\mathbf{x},\pi(\mathbf{x}))}{\vartheta_{f_{\boldsymbol{\omega}}}(Y^Q\,;\,\mathbf{x},\pi(\mathbf{x}))}\right)\big|\mathbf{X}^Q=\mathbf{x}\right] q(\mathbf{x})\,d\mathbf{x} \\
&\overset{(ii)}{\le} \sum_{j\in[m]:\omega_j'\neq\omega_j} Q\big(B_{r/4}(\mathbf{x}_j^*)\big)\,\mathbb{E}_{\boldsymbol{\omega}'}[I_{r,\pi}^Q(\tilde{p})]\,\mathrm{KL}(\text{Bernoulli}(1/2+C_\beta r^\beta)\|\text{Bernoulli}(1/2-C_\beta r^\beta)) \\
&\overset{(iii)}{\le} 32C_\beta^2 r^{2\beta} \sum_{j\in[m]:\omega_j'\neq\omega_j} Q\big(B_{r/4}(\mathbf{x}_j^*)\big)\mathbb{E}_Q[I_{r,\pi}^Q(\tilde{p})] \\
&\overset{(iv)}{\le} 32C_\beta^2\frac{1}{\lfloor 1/r\rfloor} m r^{d+2\beta} \\
&\overset{(v)}{\lesssim} r^{d+1+2\beta} m,
\end{aligned}$$

where

- step (i) uses the cancellation inside the log due to Lemma 13, and the construction of $q(\mathbf{x})$ given in (22), which is zero on the annulus between the r-ball and the $r/4$-ball,
- step (ii) uses the definition in (27) and the construction of the reward function given in (29),
- step (iii) uses Lemma 5,
- step (iv) uses the construction of $q(\mathbf{x})$ given in (22), which is equal to $q_1 = r^d/\mathrm{Leb}\big(B_{r/4}(\mathbf{x}_j^*)\big)$ inside the $r/4$-balls, and the inequality (28),
- step (iv) uses $1/\lfloor 1/r\rfloor \lesssim r$ and the fact that C_β does not depend on n, m or r.

This completes the proof of Lemma 20. □

E Proof of Private Upper Bound

E.1 Proof of Proposition 3

Proof. Given a target context $\mathbf{x}^Q \in \mathcal{X}$ and treatment candidate $t \in \mathcal{T}$, let us assume $(\mathbf{x}^Q, t) \in A_{h,j}$, without loss of generality. Let us define the empirical measures as

$$\mu_n(A_{h,j}) := \frac{1}{n}\sum_{i=1}^n \mathbb{1}\big((X_i^P, t_i^P)\in A_{h,j}\big), \quad \nu_n(A_{h,j}) := \frac{1}{n}\sum_{i=1}^n Y_i\,\mathbb{1}\big((X_i^P, t_i^P)\in A_{h,j}\big). \tag{43}$$

We define a binning-based analogue of the non-private pessimistic estimator introduced in Proposition 1:

$$\tilde{f}_{h,\mathrm{BIN}}(\mathbf{x}^Q,t) := \left(\frac{\nu_n(A_{h,j})}{\mu_n(A_{h,j})} - \sqrt{\frac{(2\log 2n)/n}{\mu_n(A_{h,j})}} - Lh^\beta\right)\mathbb{1}\big(\mu_n(A_{h,j})>0\big). \tag{44}$$

Following the proof strategy in Appendix C, it is straightforward to show that, with probability at least $1-1/n$,

$$\tilde{f}_{h,\mathrm{BIN}}(\mathbf{x}^Q,t) \le f(\mathbf{x}^Q,t).$$

Therefore, it suffices to show that, with probability at least $1-3/n$,

$$\tilde{f}_{h,\mathrm{DP}}(\mathbf{x}^Q,t) \le \tilde{f}_{h,\mathrm{BIN}}(\mathbf{x}^Q,t).$$

Since $\xi_{i,j}$'s and $\zeta'_{i,j}s$ are are i.i.d. standard Laplace scaled by $4\sqrt{2}/\varepsilon$. Each variable is sub-exponential with parameters $(\nu, \alpha) = (8\sqrt{2}/\varepsilon, 8\sqrt{2}/\varepsilon)$, so their average over n terms is sub-exponential with parameters $(\nu, \alpha) = (8\sqrt{2}/(\varepsilon\sqrt{n}), 8\sqrt{2}/(\varepsilon n))$. Therefore, by sub-exponential tail bound, we have

$$\tilde{\nu}_n(A_{h,j}) - \nu_n(A_{h,j}) < t_\varepsilon \text{ with probability at least } 1 - 1/n, \tag{45}$$

$$\tilde{\mu}_n(A_{h,j}) - \mu_n(A_{h,j}) < t_\varepsilon \text{ with probability at least } 1 - 1/n, \tag{46}$$

$$\mu_n(A_{h,j}) - \tilde{\mu}_n(A_{h,j}) < t_\varepsilon \text{ with probability at least } 1 - 1/n, \tag{47}$$

where we recall

$$t_\varepsilon = \frac{16}{\varepsilon}\sqrt{\frac{\log n}{n}}.$$

The case of $\mu_n(A_{h,j}) = 0$. In this case, we have $\tilde{f}_{h,\mathrm{BIN}}(\mathbf{x}^Q, t) = 0$. We also have, with probability at least $1 - 1/n$,

$$\mu_n = 0 \iff \tilde{\mu}_n - \mu_n = \tilde{\mu}_n \overset{(i)}{\Longrightarrow} \tilde{\mu}_n < \frac{16}{\varepsilon}\sqrt{\frac{\log n}{n}} \implies \tilde{f}_{h,\mathrm{DP}}(\mathbf{x}^Q, t) = 0,$$

where step (i) uses (46). Since the range of the reward function is $[0, 1]$, we have, with probability at least $1 - 1/n$,

$$\tilde{f}_{h,\mathrm{DP}}(\mathbf{x}^Q, t) \le \tilde{f}_{h,\mathrm{BIN}}(\mathbf{x}^Q, t) \le f(\mathbf{x}^Q, t).$$

The case of $\mu_n(A_{h,j}) > 0$ *and* $\tilde{\mu}_n < t_\varepsilon$ In this case, we have $\tilde{f}_{h,\mathrm{DP}}(\mathbf{x}^Q, t) = 0$. Since the range of the reward function is $[0, 1]$, we have, almost surely,

$$\tilde{f}_{h,\mathrm{DP}}(\mathbf{x}^Q, t) \le f(\mathbf{x}^Q, t).$$

The case of $\mu_n(A_{h,j}) > 0$ *and* $\tilde{\mu}_n > t_\varepsilon$ In this case, it suffices to show that, with probability at least $1 - 3/n$,

$$\frac{\tilde{\nu}_n(A_{h,j}) - t_\varepsilon}{\tilde{\mu}_n(A_{h,j}) + t_\varepsilon} - \sqrt{\frac{(2\log 2n)/n}{\tilde{\mu}_n(A_{h,j}) - t_\varepsilon}} \le \frac{\nu_n(A_{h,j})}{\mu_n(A_{h,j})} - \sqrt{\frac{(2\log 2n)/n}{\mu_n(A_{h,j})}}.$$

To this end, it suffices to show that, with probability at least $1 - 3/n$,

$$\begin{aligned} \tilde{\nu}_n(A_{h,j}) - t_\varepsilon &\le \nu_n(A_{h,j}), \\ \tilde{\mu}_n(A_{h,j}) + t_\varepsilon &\ge \mu_n(A_{h,j}), \\ \tilde{\mu}_n(A_{h,j}) - t_\varepsilon &\le \mu_n(A_{h,j}), \end{aligned}$$

which are equivalent to (45), (47), and (46). This completes the proof of Proposition 3. □