Minimax estimation of optimal treatment values with discrete covariates

CausalSmith*

September 10, 2026

Abstract

This paper characterizes the finite-sample minimax risk for estimating the scalar optimal-treatment value with a binary treatment, binary outcome, and a discrete covariate alphabet of size d. Under consistency, conditional exchangeability, i.i.d. observed sampling, and a fixed overlap parameter $\epsilon \in (0, 1/2)$, the causal value equals the observed-law functional that sums, over covariate cells, the larger of the two arm-specific outcome regressions. For sample size n, the fixed-sample minimax squared risk satisfies

$$\mathfrak{R}_{n,d,\epsilon} \asymp_\epsilon \min\left\{1, \frac{d}{n\log(ed)}\right\}.$$

The upper bound is attained by an explicit Jackson factorial estimator: pilot counts form local rectangles for each four-coordinate treatment–outcome cell table, a tensor Jackson polynomial approximates the cellwise maximum contribution, centered factorial moments lift the polynomial into unbiased count statistics under Poisson splitting, and Rao–Blackwellization returns an all-data fixed-sample estimator. The lower bound embeds normalized two-sample L_1 distance into an equal-propensity slice of the observed model and uses moment matching for the absolute-value cusp. The matched rate gives uniform consistency exactly when $d = o\{n \log(en)\}$ and gives the parametric n^{-1} minimax scale exactly along bounded-alphabet sequences.

1 Introduction

Optimal treatment choice assigns the arm with the larger conditional outcome mean to each covariate profile. In a discrete covariate model, the population value of this oracle rule is a scalar obtained by summing cell masses times the larger of two arm-specific regressions. This scalar is central in treatment-choice analysis because it reports the welfare level achieved by the unrestricted cellwise oracle assignment, the benchmark against which any particular learned rule is measured. The statistical difficulty comes from estimating many cellwise maxima when the covariate alphabet grows with the sample size.

This paper studies that difficulty under the fixed-overlap observed experiment. The data are i.i.d. triples $O = (X, A, Y)$, where $X \in [d]$, $A \in \{0, 1\}$, and $Y \in \{0, 1\}$. The observed law belongs

*Machine-generated by the CausalSmith research pipeline. Each displayed formal statement is either mapped to a Lean declaration or marked as a presentation-level definition, and each displayed proof renders a Lean proof, as recorded in the verification appendix (scope, toolchain, commit, and any statement conditional on a published input). Cited work otherwise supplies framing and positioning.

to $\mathcal{D}^{\text{obs}}_{d,\epsilon}$, the fixed-overlap class defined in Definition 3, with overlap parameter $\epsilon \in (0, 1/2)$. The target is the observed-law optimal-regression value $\Psi(\mathbb{P})$ from Definition 4. Under the full-data completion class $\mathcal{P}_{d,\epsilon}$ in Definition 6, which imposes consistency, conditional exchangeability, and the same fixed-overlap observed margin, Proposition 2 identifies this observed value with the causal oracle value $V^\star(P)$ from Definition 9.

The main result is a matched all-regime finite-sample minimax law. With $L_d = \log(ed)$, Theorem 3 proves that, for every fixed $\epsilon \in (0, 1/2)$, there are constants $0 < c_\epsilon \le C_\epsilon < \infty$ such that for every $n \ge 1$ and $d \ge 2$,

$$c_\epsilon \min\left\{1, \frac{d}{nL_d}\right\} \le \mathfrak{R}_{n,d,\epsilon} \le C_\epsilon \min\left\{1, \frac{d}{nL_d}\right\},$$

where $\mathfrak{R}_{n,d,\epsilon}$, the fixed-sample observed minimax squared risk, is defined in Definition 5. The same theorem states that the corresponding minimax risk over the causal completion class equals this observed-law risk. Thus the scalar oracle value is uniformly estimable precisely up to the rate dictated by the alphabet-to-sample ratio $d/\{n \log(ed)\}$.

The estimator achieving the upper bound is written out in closed form. The empirical-ratio estimator in Definition 10 serves as the bounded-alphabet branch. For growing alphabets, Algorithm 1 defines a Jackson factorial estimator that splits a Poissonized subsample into pilot and evaluation counts, builds a pilot-local rectangle for each covariate cell, approximates the globally defined cell contribution $\bar{f}_\epsilon$ from Definition 8 by a tensor Jackson polynomial, replaces normalized monomials by centered factorial lifts, clips each cell estimate at its pilot scale, projects the aggregate to $[0, 1]$, and Rao–Blackwellizes over the auxiliary randomization. Theorem 1 gives the resulting uniform risk bound over $\mathcal{D}^{\text{obs}}_{d,\epsilon}$. The construction is stated as a rate-achieving decision-theoretic procedure: every ingredient is defined by an explicit formula, yielding a complete statistical specification of the tensor-polynomial estimator and Rao–Blackwellized rule at the stated alphabet sizes and approximation orders.

In the proof, the logarithmic factor is tied to the absolute-value/L_1 lower-bound mechanism for the cellwise maximum. On the overlap cone of Definition 7, the cell contribution equals the larger arm-specific regression contribution; after normalization, its hard local component has the geometry of an absolute value. The polynomial approximation and factorial-moment architecture follows the large-alphabet nonsmooth-functional literature, where approximation of $|t|$, moment matching, and unbiased polynomial lifts generate effective logarithmic sample-size gains [Cai and Low, 2011, Jiao et al., 2015, Wu and Yang, 2016, Han et al., 2018, Valiant and Valiant, 2017]. The tensor Jackson certificate in Lemma 4 supplies the local approximation and coefficient control needed for the four-coordinate treatment–outcome tables.

The converse embeds a classical large-alphabet testing problem into the treatment-value model. Proposition 3 maps two distributions $(\mathbf{P}, \mathbf{Q}) \in \Delta_d^2$ into an equal-propensity observed law whose optimal-regression value is $1/2 + \|\mathbf{P} - \mathbf{Q}\|_1/4$. This exact Markov reduction transfers the normalized two-sample L_1 minimax obstruction of Jiao et al. [2018] to the observed optimal-value problem. In the dense regime, Lemma 10 gives the corresponding moment-matched construction using the absolute-value approximation lower bound of Cai and Low [2011]. Theorem 2 combines these ingredients into the all-estimator lower bound matching the Jackson factorial upper rate.

The rate has direct implications for growing categorical adjustment. Theorem 4 shows that, along alphabet sequences d_n, the observed and causal minimax risks converge to zero exactly when $d_n = o\{n \log(en)\}$. The same theorem characterizes the parametric n^{-1} minimax scale exactly by $d_n = O(1)$. These statements translate the matched finite-sample frontier into operational growth rules for discrete covariate tables under fixed overlap.

The analysis connects treatment-choice econometrics with nonregular and large-alphabet estimation. Treatment-choice and welfare-regret work studies optimal rules and policy performance [Manski, 2004, Murphy, 2003, Hirano and Porter, 2009, Kitagawa and Tetenov, 2018, Qian and Murphy, 2011, Athey and Wager, 2021]. Semiparametric analyses of optimal values describe identification, efficient influence functions, smoothing, and fixed-law inference under regularity or local conditions [Robins, 1986, Rosenbaum and Rubin, 1983, van der Laan and Luedtke, 2015, Luedtke and van der Laan, 2016, Chakraborty et al., 2010, Hirano and Porter, 2012, Chen et al., 2023, Whitehouse et al., 2025, Wei, 2025]. The present result gives a finite-sample minimax characterization for the scalar unrestricted cellwise oracle value over a growing discrete alphabet, with fixed overlap and arbitrary cell masses, null cells, boundary binary means, and treatment-effect ties inside the model class.

A verification appendix records the verification status and the cited external inputs for the anchored statements. The paper proceeds by reviewing related work, defining the observed and causal model classes, stating the estimator and minimax theorems, discussing implications and the sharp-constant question, and then giving the analytic and lower-bound details in the appendices.

2 Related work

This paper belongs to the treatment-choice literature that studies how observational or experimental data can support decisions that assign treatment by covariate profile. The classical welfare and empirical-welfare strands analyze treatment rules, regret, and policy choice when the object of interest is a decision rule or its welfare performance [Manski, 2004, Murphy, 2003, Hirano and Porter, 2009, Kitagawa and Tetenov, 2018, Qian and Murphy, 2011, Athey and Wager, 2021]. The present analysis focuses on the scalar value attained by choosing the better arm within each discrete covariate cell. That focus isolates the nonsmooth cellwise maximum that enters the optimal value and makes the alphabet size of the covariate distribution a primitive dimension of the estimation problem.

A second connection is to causal identification and semiparametric estimation of optimal values. Identification under consistency, conditional exchangeability, and overlap traces to the potential-outcome and inverse-probability traditions [Robins, 1986, Rosenbaum and Rubin, 1983]. Targeted-learning and efficient-influence-function analyses then study optimal-value parameters and data-adaptive treatment rules under regularity conditions that support asymptotic distribution theory [van der Laan and Luedtke, 2015, Luedtke and van der Laan, 2016]. The fixed-overlap discrete model considered here keeps the same causal target once the observed law identifies it, and then asks for the finite-sample minimax risk of estimating the identified scalar optimal value.

The cellwise maximum also places the problem in the nonregular-inference literature. Optimal treatment values have kink points where two treatment arms have equal conditional means, and those ties create the familiar failure of ordinary differentiability for max-type functionals [Chakraborty et al., 2010, Hirano and Porter, 2012]. Recent work develops smoothing, debiasing, and local-asymptotic analyses for related nonregular value and policy-learning targets [Chen et al., 2023, Whitehouse et al., 2025, Wei, 2025]. The contribution here characterizes the minimax scale for the discrete-covariate scalar value under fixed overlap, with the dense-alphabet regime governed by approximation of the absolute-value singularity induced by the maximum.

That approximation structure links the analysis to large-alphabet estimation of nonsmooth distributional functionals. Polynomial approximation and moment matching have proved central for sharp rates in estimating entropy, support-related quantities, and distances between high-dimensional discrete distributions [Jiao et al., 2015, Wu and Yang, 2016, Han et al., 2018, Valiant and Valiant, 2017], while Cai and Low [2011] supplies the sharp Gaussian absolute-value-cusp analysis used in the lower-bound comparison. The closest analogue is the normalized two-sample L_1 distance problem of Jiao et al. [2018], where the absolute-value functional yields logarithmic factors through best polynomial approximation. In the treatment-value problem, fixed overlap maps each covariate cell to a four-atom observed table and the maximum over arms creates the same absolute-value geometry after normalization.

Two large-alphabet papers are close enough to compare rate by rate. Zeng et al. [2024] study causal inference with unrestricted high-dimensional discrete covariates. Their target is a linear treatment-specific mean under an alphabet of size d, estimated from n i.i.d. observations; they show that regression, weighting, and doubly robust estimators attain mean-squared error of order $d^2/n^2 + 1/n$, prove a minimax lower bound of order $d^2/\{n^2 \log^2 n\} + 1/n$, and obtain the faster order $d/n^2 + 1/n$ under effect homogeneity or a known covariate distribution. Their aggregation is signed and linear, so the leading difficulty is bias accumulation across cells. The target of the present paper replaces that signed sum by the cellwise maximum, and the resulting frontier is $\min\{1, d/[n \log(ed)]\}$: a different exponent in d, a different consistency boundary, and a logarithmic gain whose source is the polynomial approximation of a cusp.

Jiao et al. [2018] give the second comparison. For two unknown d-point distributions observed through independent Poissonized samples of size $\mathrm{Pois}(2n)$ each, they construct minimax-rate estimators of $\|\mathbf{P} - \mathbf{Q}\|_1$ and prove the matching $d/\{n \log(en)\}$ squared-error obstruction, so their result is a full estimator-and-converse analysis of the two-sample L_1 distance. The present lower bound transfers that converse into the treatment-value model through the equal-propensity embedding of Proposition 3, and the upper bound is constructed here for the observational cell table. Three features of the observational problem require new work. The arm denominators are unknown and are estimated from the same sample, whereas the two-sample problem fixes the sample sizes by design; the propensities are unequal and constrained only by fixed overlap, so the two arms of a cell carry different effective sample sizes; and each covariate cell is a grouped four-atom table, where the two-sample problem presents a pair of scalar masses, which is why the upper bound needs a four-coordinate tensor Jackson construction with pilot-adaptive rectangles.

Standard minimax tools such as two-point and moment-matching lower bounds, together with approximation-theoretic upper bounds [Le Cam, 1986, Tsybakov, 2009, DeVore and Lorentz, 1993], provide the methodological basis for matching the estimator and lower bound in this setting.

3 Setup and assumptions

Throughout, $d \geq 2$ denotes the covariate alphabet size, n the fixed sample size, and $\epsilon \in (0, 1/2)$ the overlap constant. Sums over covariate cells run over $[d]$, and treatment and outcome indices take values in $\{0, 1\}$. Generic positive finite constants may change from line to line; constants indexed by ϵ may depend on the fixed overlap constant. Vector norms are the usual finite-dimensional norms, with ℓ_1 written explicitly when it is used.

We begin with the elementary projection notation used later by the estimators.

Definition 1. [`synth_13`] (Interval projection) *For real numbers $a \leq b$, let $\Pi_{[a,b]} : \mathbb{R} \to [a, b]$ be the interval projection $\Pi_{[a,b]}(z) = \min\{b, \max\{a, z\}\}$.*

The statistical experiment is a finite observational study. One unit is the observed triple O_i=(X_i,A_i,Y_i), where the covariate lies in the finite alphabet $[d]$ and the treatment and outcome are binary. The observed law $\mathbb{P}$ is a probability law on this finite sample space.

Assumption 1. [`ass:iid-sampling`] (I.i.d. observed sampling) *For each observed law $\mathbb{P} \in \mathcal{D}^{\mathrm{obs}}_{d,\epsilon}$, let $O_i = (X_i, A_i, Y_i)$ denote the observed triple for unit i. The observations $O_1, \dots, O_n$ are independent and identically distributed with common law $\mathbb{P}$.*

Assumption 1 is the standard independent and identically distributed sampling condition for the observed data [Zeng et al., 2024]. It fixes the repeated-sampling experiment under each admissible observed law and makes the minimax risk below a fixed-sample decision problem.

For causal interpretation, we place the observed experiment inside a finite binary potential-outcome model. A full-data law P governs $(X, A, Y, Y(0), Y(1))$ on the same covariate alphabet.

Assumption 2. [`ass:consistency`] (Consistency of outcomes) *The observed outcome satisfies*

$$Y = Y(A)$$

almost surely.

Assumption 2 is the standard consistency condition [Zeng et al., 2024]. It ties the realized outcome to the potential outcome under the realized treatment arm, so the observed binary response carries the arm-specific potential-outcome information in treated and untreated cells.

Assumption 3. [`ass:conditional-exchangeability`] (Conditional exchangeability) *For each $a \in \{0, 1\}$, the potential outcome $Y(a)$ is conditionally independent of treatment given covariates:*

$$Y(a) \perp A \mid X.$$

Assumption 3 is the standard armwise conditional exchangeability condition [Zeng et al., 2024]. It encodes the observational-design requirement that, within each covariate cell, treatment assignment carries no further information about each potential outcome.

Assumption 4. [`ass:fixed-overlap`] (Fixed overlap) *Let $[d] = \{1, \dots, d\}$ be the finite covariate alphabet. For every observed law $\mathbb{P}$ under consideration and every $x \in [d]$ with covariate-cell mass $p_x = \mathbb{P}(X = x) > 0$, the treatment propensity $\pi_x = \mathbb{P}(A = 1 \mid X = x)$ satisfies*

$$\epsilon \le \pi_x \le 1 - \epsilon,$$

where $\epsilon \in (0, 1/2)$ is the fixed overlap constant.

Assumption 4 is the standard fixed strong-overlap condition [Zeng et al., 2024]. The propensity π_x is bounded away from the edges on every occupied covariate cell x, with p_x denoting the corresponding cell mass; this keeps both treatment arms represented at the population level in each relevant cell.

We next define the observed-law model and the target functional. The observed margin records exactly the distribution of the observable triple induced by a full-data law.

Definition 2. [`def:observed-margin`] (Observed margin $\mathrm{Obs}(P)$) *For a full-data law P, its observed margin $\mathrm{Obs}(P)$ is the law on $[d] \times \{0, 1\}^2$ defined by*

$$\mathrm{Obs}(P)\{(x, a, y)\} = P(X = x, A = a, Y = y), \qquad (x, a, y) \in [d] \times \{0, 1\}^2.$$

The observed class parameterizes each covariate cell by its mass, propensity, and arm-specific binary outcome mean μ_{ax}. Its atom probabilities $q_{ay,x}$ are arranged into the cell vector q_x, and the table across cells is denoted by $\mathbf{q}$.

Definition 3. [`def:observed-model-class`] (Observed-law class $\mathcal{D}_{d,\epsilon}^{\text{obs}}$) *Under Assumption 4, let*

$$\Delta_d = \left\{ p \in [0,1]^d : \sum_{x=1}^{d} p_x = 1 \right\}.$$

The class $\mathcal{D}_{d,\epsilon}^{\text{obs}}$ consists of all laws $\mathbb{P}$ on $[d] \times \{0,1\}^2$ for which there exist $p = (p_1, \ldots, p_d) \in \Delta_d$, propensities $\pi_x \in [\epsilon, 1-\epsilon]$, and means $\mu_{ax} \in [0,1]$ such that, for every $x \in [d]$ and $a \in \{0,1\}$,

$$q_{a1,x} = p_x \pi_x^a (1-\pi_x)^{1-a} \mu_{ax}, \qquad q_{a0,x} = p_x \pi_x^a (1-\pi_x)^{1-a} (1-\mu_{ax}),$$

where

$$q_{ay,x} = \mathbb{P}(X = x, A = a, Y = y).$$

For each cell x, write

$$q_x = (q_{00,x}, q_{01,x}, q_{10,x}, q_{11,x}).$$

The support contains at most d covariate cells, and on any null cell the parameters may be chosen arbitrarily within the displayed ranges.

Within this observed class, the estimand is the population value obtained by choosing the better arm in each covariate cell according to the conditional outcome means.

Definition 4. [`def:observed-optimal-value`] (Observed-law optimal value $\Psi(\mathbb{P})$) *For $\mathbb{P} \in \mathcal{D}_{d,\epsilon}^{\text{obs}}$ as in Definition 3, with table $\mathbf{q} = (q_x)_{x=1}^d$, the observed-law optimal-regression functional is*

$$\Psi(\mathbb{P}) = \Psi(\mathbf{q}) := \sum_{x=1}^{d} p_x \max_{a \in \{0,1\}} \mu_{ax}.$$

This value is independent of the permitted choice of μ_{ax} on cells with $p_x = 0$.

The risk criterion evaluates estimators of $\Psi(\mathbb{P})$ uniformly over the fixed-overlap observed laws.

Definition 5. [`def:minimax-risk`] (Minimax risk $\mathfrak{R}_{n,d,\epsilon}$) *The fixed-sample minimax squared risk is*

$$\mathfrak{R}_{n,d,\epsilon} := \inf_{\widehat{V}} \sup_{\mathbb{P} \in \mathcal{D}_{d,\epsilon}^{\text{obs}}} E_{\mathbb{P}} \left[\left\{ \widehat{V}(O_1, \ldots, O_n) - \Psi(\mathbb{P}) \right\}^2 \right],$$

where the infimum is over all measurable maps

$$\widehat{V} : \left([d] \times \{0,1\}^2 \right)^n \to \mathbb{R}.$$

The full-data causal class collects exactly the potential-outcome laws whose observed margins belong to the preceding observed-law model and whose potential outcomes satisfy the causal conditions already stated.

Definition 6. [`def:model-class`] (Full-data causal class $\mathcal{P}_{d,\epsilon}$) *For every $d \geq 2$ and $0 < \epsilon < 1/2$, $\mathcal{P}_{d,\epsilon}$ is the set of probability laws P for $(X, A, Y, Y(0), Y(1))$ such that:*

- ***(Support.)*** *$X \in [d]$ and $A, Y, Y(0), Y(1) \in \{0,1\}$ P-almost surely.*
- ***(Potential-outcome structure.)*** *P satisfies Assumptions 2 and 3.*
- ***(Observed fixed-overlap margin.)*** *The observed margin $\mathrm{Obs}(P)$ from Definition 2 belongs to the fixed-overlap observed-law class $\mathcal{D}^{\mathrm{obs}}_{d,\epsilon}$ under Assumption 4.*

The analytic representation of the target works cell by cell. For a generic four-vector u, the next definition records the overlap geometry in terms of arm masses $s_a(u)$ and total mass $s(u)$.

Definition 7. [`def:overlap-cone`] (Overlap cone $\mathcal{C}_\epsilon$) *The four-cell overlap cone is*

$$\mathcal{C}_\epsilon := \{u \in \mathbb{R}^4_+ : \epsilon s(u) \le s_1(u) \le (1-\epsilon)s(u)\},$$

where, for each $a \in \{0,1\}$,

$$s_a(u) = u_{a0} + u_{a1}, \qquad s(u) = s_0(u) + s_1(u).$$

The cell contribution is then extended from the overlap region to the whole nonnegative orthant by stabilizing the arm denominators.

Definition 8. [`def:global-extension`] (Global extension $g_{a,\epsilon}$ and $\bar{f}_\epsilon$) *For $u \in \mathbb{R}^4_+$ with $s(u) > 0$, define the overlap-stabilized arm contribution*

$$g_{a,\epsilon}(u) := \frac{s(u)u_{a1}}{\max\{s_a(u), \epsilon s(u)\}}, \qquad a \in \{0,1\},$$

and the globally defined cellwise maximum contribution

$$\bar{f}_\epsilon(u) := \max_{a\in\{0,1\}} g_{a,\epsilon}(u).$$

At the origin,

$$g_{a,\epsilon}(0) = 0, \qquad \bar{f}_\epsilon(0) = 0.$$

Finally, the causal optimal-treatment value is the full-data analogue of the observed optimal-regression value.

Definition 9. [`def:oracle-value`] (Oracle value $V^\star(P)$) *For a full-data law P, the full-data causal optimal-treatment value is*

$$V^\star(P) = E_P\left[\max_{a\in\{0,1\}} E_P\{Y(a) \mid X\}\right].$$

The following result connects the observed-law representation, the analytic extension, and the causal completion class.

Proposition 1. [`prop:identification-and-extension`] (Identification and completion) *Let $d \ge 2$, let $\epsilon \in (0, 1/2)$, and let $\mathbb{P} \in \mathcal{D}^{\mathrm{obs}}_{d,\epsilon}$ in the sense of Definition 3. Then:*

- ***(Cell overlap.)*** *For every $x \in [d]$, the four-vector $q_x = (q_{00,x}, q_{01,x}, q_{10,x}, q_{11,x})$ belongs to the overlap cone $\mathcal{C}_\epsilon$ of Definition 7.*

- ***(Observed value.)*** *The observed-law optimal-regression value of Definition 4 satisfies*

$$\Psi(\mathbb{P}) = \sum_{x=1}^{d} \bar{f}_\epsilon(q_x).$$

- ***(Global extension bounds.)*** *For every $u \in \mathbb{R}_+^4$, the global extension of Definition 8 satisfies*

$$0 \le \bar{f}_\epsilon(u) \le s(u).$$

- ***(Global Lipschitz property.)*** *For every $u, v \in \mathbb{R}_+^4$,*

$$\left|\bar{f}_\epsilon(u) - \bar{f}_\epsilon(v)\right| \le (1 + \epsilon^{-1}) \sum_{\jmath} |u_\jmath - v_\jmath| \,.$$

- ***(Completion.)*** *There exists a full-data law $P \in \mathcal{P}_{d,\epsilon}$ in the sense of Definition 6 whose observed margin satisfies $\mathrm{Obs}(P) = \mathbb{P}$. The completion is consistent and conditionally exchangeable, treatment is independent of the joint potential-outcome vector given X, and the two potential outcomes are conditionally independent Bernoulli variables given X.*

- ***(Observed margins.)*** *For every full-data law $P \in \mathcal{P}_{d,\epsilon}$, its observed margin $\mathrm{Obs}(P)$ of Definition 2 belongs to $\mathcal{D}_{d,\epsilon}^{\mathrm{obs}}$.*

Proposition 1 establishes three facts used throughout the analysis. First, every observed cell lies in the overlap cone. Second, the observed value decomposes as the sum of the globally defined cell contributions $\bar{f}_\epsilon$. Third, the extension is bounded and Lipschitz on the nonnegative orthant, which supplies the analytic control used by the estimator construction in the main results.

The completion clauses give the causal interpretation of the observed model. Every observed law in $\mathcal{D}_{d,\epsilon}^{\mathrm{obs}}$ admits a full-data law in $\mathcal{P}_{d,\epsilon}$ with the same observed margin, and every law in $\mathcal{P}_{d,\epsilon}$ returns an observed margin in $\mathcal{D}_{d,\epsilon}^{\mathrm{obs}}$. Thus the observed minimax problem is aligned with a potential-outcome class satisfying consistency, conditional exchangeability, and fixed overlap.

Proposition 2. [`prop:causal-optimal-value-corollary`] (Causal value identification) *Let $d \ge 2$, let $\epsilon \in (0, 1/2)$, and let P be a full-data law satisfying $\mathcal{P}_{d,\epsilon}$ in the sense of Definition 6. Then the full-data oracle value in Definition 9 agrees with the observed-law optimal-regression value in Definition 4 evaluated at the observed margin in Definition 2:*

$$V^\star(P) = \Psi\{\mathrm{Obs}(P)\}.$$

Proposition 2 identifies the full-data optimal-treatment value $V^\star(P)$ with the observed-law functional evaluated at the observed margin $\mathrm{Obs}(P)$. The equality lets the minimax analysis target Ψ on $\mathcal{D}_{d,\epsilon}^{\mathrm{obs}}$ while retaining its causal interpretation for laws in $\mathcal{P}_{d,\epsilon}$.

4 Main results

The setup in Definitions 3 to 5 turns the value problem into uniform estimation of the observed-law optimal-regression value over the fixed-overlap observed model. This section gives estimators and finite-sample risk bounds for that problem. The final statements then transfer the same rate to the causal completion class through the identification result in Proposition 2.

For a fixed alphabet, the natural benchmark is the empirical cell-ratio estimator $\widehat{V}^{\mathrm{ER}}_{n,d}$. It estimates each covariate-cell mass and each arm-specific success probability by its sample analogue, then maximizes over treatment arms cell by cell.

Definition 10. [`def:empirical-ratio-estimator`] (Empirical-ratio estimator $\widehat{V}^{\mathrm{ER}}_{n,d}$) *For $x \in [d]$ and $a \in \{0,1\}$, define the empirical counts*

$$\overline{N}_x = \sum_{i=1}^{n} \mathbf{1}\{X_i = x\}, \qquad \overline{N}_{ax} = \sum_{i=1}^{n} \mathbf{1}\{X_i = x, A_i = a\},$$

and

$$\overline{S}_{ax} = \sum_{i=1}^{n} \mathbf{1}\{X_i = x, A_i = a, Y_i = 1\}.$$

Set

$$\widehat{p}_x = \frac{\overline{N}_x}{n}, \qquad \widehat{\mu}_{ax} = \frac{\overline{S}_{ax}}{\overline{N}_{ax}},$$

with the convention $0/0 = 0$. The empirical-ratio optimal-value estimator is

$$\widehat{V}^{\mathrm{ER}}_{n,d} = \sum_{x=1}^{d} \widehat{p}_x \max_{a \in \{0,1\}} \widehat{\mu}_{ax}.$$

This plug-in form is useful at bounded alphabet sizes. The high-dimensional construction below instead estimates the cell contribution through a polynomial approximation to the Lipschitz extension $\bar{f}_\epsilon$ from Definition 8. The construction follows the standard use of polynomial approximation for nonsmooth functionals [DeVore and Lorentz, 1993, Jiao et al., 2015].

Definition 11. [`synth_18`] (Treatment-outcome coordinate set) *The treatment-outcome coordinate set $\mathcal{J}$ is*

$$\mathcal{J} := \{0,1\} \times \{0,1\}.$$

A coordinate index is denoted by $\jmath \in \mathcal{J}$.

The set $\mathcal{J}$ indexes the four treatment-outcome atom coordinates in each covariate cell. Given a pilot rectangle in these four coordinates, the next definition records the Jackson polynomial used to approximate the cellwise contribution.

Definition 12. [`def:jackson-kernel`] (Jackson polynomial P_{x,K_d}) *For an integer $K_d \geq 2$, define the order-four Jackson kernel on $[-\pi, \pi]$ by*

$$J_{K_d}(t) = c_{K_d} \left\{ \frac{\sin(K_d t/2)}{\sin(t/2)} \right\}^4,$$

with removable values filled continuously and normalizing constant c_{K_d} chosen so that

$$\int_{-\pi}^{\pi} J_{K_d}(t)\, dt = 1.$$

For a rectangle

$$Q_x = \prod_{\jmath \in \mathcal{J}} [\ell_{\jmath,x}, u_{\jmath,x}],$$

define its coordinate centers and radii by

$$b_{\jmath,x} = \frac{\ell_{\jmath,x} + u_{\jmath,x}}{2}, \qquad r_{\jmath,x} = \frac{u_{\jmath,x} - \ell_{\jmath,x}}{2}.$$

The associated tensor polynomial P_{x,K_d} is defined by the even product convolution

$$P_{x,K_d}(b_x + r_x \odot \cos\theta) = \int_{[-\pi,\pi]^4} \bar{f}_\epsilon\{b_x + r_x \odot \cos(\theta + t)\} \prod_{\jmath \in \mathcal{J}} J_{K_d}(t_\jmath)\, dt.$$

The order K_d grows on the logarithmic alphabet scale, so the approximation error and coefficient growth can be balanced against sampling noise. In the algorithm, $\Pi_{[a,b]}$ denotes projection onto the interval $[a,b]$, and the constants H_0, κ, and D_0 are fixed tuning constants chosen in the risk theorem.

Algorithm 1 (Jackson-factorial estimator $\widehat{V}^{\mathrm{JF}}_{n,d}$)

The inputs are $O_1, \ldots, O_n$, the alphabet size d, and tuning constants D_0, κ, and H_0.

1. Define the logarithmic scale, Poissonized scale, and polynomial degree by

$$L_d = \log(ed), \qquad m = \frac{n}{8}, \qquad K_d = \max\{2, \lfloor \kappa L_d \rfloor\}.$$

2. If $d < D_0$, output

$$\widehat{V}^{\mathrm{JF}}_{n,d} = \widehat{V}^{\mathrm{ER}}_{n,d}.$$

If $d \geq D_0$ and $n < d/L_d$, output

$$\widehat{V}^{\mathrm{JF}}_{n,d} = \frac{1}{2}.$$

3. In the remaining case, draw

$$M \sim \mathrm{Pois}(n/4).$$

On the event $M > n$, set

$$\widetilde{V} = 0.$$

4. On the event $M \leq n$, draw independent marks

$$B_i \sim \mathrm{Bernoulli}(1/2), \qquad i = 1, \ldots, M.$$

For each $x \in [d]$ and coordinate $\jmath \in \mathcal{J}$, define pilot and evaluation counts

$$N'_{\jmath,x} = \sum_{i=1}^{M} B_i\, \mathbf{1}\{O_i \text{ belongs to coordinate } \jmath \text{ in cell } x\},$$

and

$$N_{\jmath,x} = \sum_{i=1}^{M} (1 - B_i)\, \mathbf{1}\{O_i \text{ belongs to coordinate } \jmath \text{ in cell } x\}.$$

5. Define pilot frequencies and rectangle half-widths by

$$c_{\jmath,x} = \frac{N'_{\jmath,x}}{m}, \qquad h_{\jmath,x} = H_0 \left\{ \sqrt{\frac{c_{\jmath,x} L_d}{m}} + \frac{L_d}{m} \right\}.$$

Set

$$\ell_{\jmath,x} = \max\{0, c_{\jmath,x} - h_{\jmath,x}\}, \qquad u_{\jmath,x} = c_{\jmath,x} + h_{\jmath,x},$$

and

$$Q_x = \prod_{\jmath \in \mathcal{J}} [\ell_{\jmath,x}, u_{\jmath,x}].$$

Let b_x and r_x denote the vector of coordinate centers and coordinate radii of Q_x.

6. Expand the local polynomial as

$$P_{x,K_d}(v) - \bar{f}_\epsilon(b_x) = \sum_\alpha b_{x,\alpha} \left(\frac{v - b_x}{r_x} \right)^\alpha.$$

For $h \geq 0$, define the centered factorial lift

$$U_h(N; z) = \sum_{t=0}^{h} \binom{h}{t} (-z)^{h-t} \frac{(N)_t}{m^t}.$$

7. Define the clipping scale

$$S_x = (1 + \epsilon^{-1}) \sum_{\jmath \in \mathcal{J}} r_{\jmath,x},$$

the unprojected cell estimate

$$Z_x = \bar{f}_\epsilon(b_x) + \sum_\alpha b_{x,\alpha} \prod_{\jmath \in \mathcal{J}} \frac{U_{\alpha_\jmath}(N_{\jmath,x}; b_{\jmath,x})}{r_{\jmath,x}^{\alpha_\jmath}},$$

and the clipped cell estimate

$$T_x^{\mathrm{JF}} = \Pi_{[\bar{f}_\epsilon(b_x) - d^{1/4} S_x,\, \bar{f}_\epsilon(b_x) + d^{1/4} S_x]}(Z_x).$$

8. Define the randomized projected aggregate by

$$\widetilde{V} = \Pi_{[0,1]} \left(\sum_{x=1}^{d} T_x^{\mathrm{JF}} \right).$$

The all-data Rao–Blackwellized estimator is

$$\widehat{V}_{n,d}^{\mathrm{JF}} = E\left[\widetilde{V} \mid O_1, \ldots, O_n \right],$$

where the conditional expectation integrates only the auxiliary Poisson count M and marks $B_1, \ldots, B_M$.

The all-data estimator $\widehat{V}_{n,d}^{\mathrm{JF}}$ uses the auxiliary Poisson count M and random split only inside a Rao–Blackwellization. Conditional on the observed sample, the final object is a deterministic measurable estimator. The centered factorial lift in the construction gives unbiased polynomial terms under the Poissonized evaluation counts, while clipping and final projection keep the aggregate on the value scale.

Definition 13. [`synth_3`] (Fixed-law squared risk) *For an estimator $\widehat{V}$ based on $O_1, \ldots, O_n$ and an observed law $\mathbb{P} \in \mathcal{D}_{d,\epsilon}^{\mathrm{obs}}$, define*

$$\mathrm{Risk}_{\mathbb{P}}(\widehat{V}, \Psi) := E_{\mathbb{P}^{\otimes n}}\left[\left\{\widehat{V}(O_1, \ldots, O_n) - \Psi(\mathbb{P})\right\}^2\right],$$

where $\mathbb{P}^{\otimes n}$ denotes the law of n independent observations each distributed as $\mathbb{P}$.

The first risk statement is uniform over the observed model class with fixed overlap. It uses the i.i.d. sampling condition from Assumption 1 and the observed-law value from Definition 4.

Theorem 1. [`thm:jackson-factorial-upper`] (Jackson factorial risk bound) *Assume the sampling condition in Assumption 1: for every alphabet size d, sample size n, and observed finite law $\mathbb{P}$, the sample law is the canonical i.i.d. product law generated by $\mathbb{P}$.*

Then there exist tuning constants $H_0 > 0$, $\kappa \in (0, 1)$, and an integer cutoff $D_0 > 2$ such that the Jackson factorial estimator $\widehat{V}_{n,d}^{\mathrm{JF}}$ of Algorithm 1 satisfies the following guarantee whenever:

- ***(Overlap.)*** *The overlap constant ϵ satisfies $0 < \epsilon < 1/2$.*
- ***(Sample size.)*** *The sample size satisfies $n \geq 1$.*
- ***(Alphabet size.)*** *The alphabet size satisfies $d \geq 2$.*
- ***(Observed model.)*** *The observed law $\mathbb{P}$ belongs to $\mathcal{D}_{d,\epsilon}^{\mathrm{obs}}$ as in Definition 3.*

For every such ϵ, there is a constant $C_\epsilon > 0$ such that, for every such n, d, and $\mathbb{P}$,

$$\mathrm{Risk}_{\mathbb{P}}\left(\widehat{V}_{n,d}^{\mathrm{JF}}, \Psi\right) \leq C_\epsilon \min\left\{1, \frac{d}{n \log(ed)}\right\}.$$

Here $\Psi(\mathbb{P})$ is the observed-law optimal-regression value of Definition 4.

Theorem 1 gives the estimator side of the minimax rate. For each fixed overlap constant, the constant C_ϵ controls the entire model class and all finite n, d. The logarithmic denominator comes from approximating the cellwise maximum by a logarithmic-degree polynomial and then estimating the resulting polynomial coefficients through factorial moments.

The converse starts from a normalized two-sample L_1 problem. Let Δ_d denote the d-point probability simplex, and let $\mathbf{P}$ and $\mathbf{Q}$ be two simplex points. The following embedding places this distance problem inside the equal-propensity slice of the observed model.

Definition 14. [`def:l1-embedding`] (L_1 embedding $(\mathbf{P}, \mathbf{Q})$) *For $(\mathbf{P}, \mathbf{Q}) \in \Delta_d^2$, define the observed atom masses*

$$q_{11,x} = q_{00,x} = \frac{P_x}{4}, \qquad q_{10,x} = q_{01,x} = \frac{Q_x}{4}, \qquad x \in [d].$$

Complete the observed law by taking the potential outcomes to be conditionally independent Bernoulli variables, setting the treatment propensity equal to $1/2$, and imposing consistency.

Definition 15. [synth_7] (Fixed-sample two-sample L_1 minimax risk) *Let $(Z_i, W_i)_{i=1}^n$ be independent paired observations with $Z_i \sim \mathbf{P}$, $W_i \sim \mathbf{Q}$, and $(\mathbf{P}, \mathbf{Q}) \in \Delta_d^2$. The normalized two-sample L_1 minimax mean-squared error is*

$$\mathfrak{R}_{n,d}^{L_1} := \inf_{\widehat{D}} \sup_{(\mathbf{P},\mathbf{Q})\in\Delta_d^2} E_{\mathbf{P},\mathbf{Q}} \left[\left\{ \widehat{D}((Z_i, W_i)_{i=1}^n) - \|\mathbf{P} - \mathbf{Q}\|_1 \right\}^2 \right],$$

where the infimum is over all measurable real-valued functions of the paired sample.

Definition 16. [synth_6] (Equal-propensity normalized L_1 embedding) *For $(\mathbf{P}, \mathbf{Q}) \in \Delta_d^2$,* $\mathrm{L1Emb}(\mathbf{P}, \mathbf{Q})$ *is the full-data law on $(X, A, Y, Y(0), Y(1))$ with covariate mass $p_x = (P_x+Q_x)/2$, treatment propensity $\pi_x = 1/2$, outcome regressions $\mu_{1x} = P_x/(P_x+Q_x)$ and $\mu_{0x} = Q_x/(P_x+Q_x)$ on cells with $P_x + Q_x > 0$, conditionally independent Bernoulli potential outcomes with these means, treatment drawn independently of the potential outcomes given X, and observed outcome $Y = Y(A)$. Equivalently, its observed margin has cell masses $q_{11,x} = q_{00,x} = P_x/4$ and $q_{10,x} = q_{01,x} = Q_x/4$, with a consistency-compatible completion.*

Under this embedding, the observed optimal value is an affine transform of the L_1 distance. The reduction is stated as an exact product-kernel comparison, in the standard experiment-comparison sense of Le Cam [1986] and Tsybakov [2009].

Proposition 3. [prop:equal-propensity-l1-reduction] (Equal-propensity L1 reduction) *Let $n, d \in \mathbb{N}$, let $\epsilon \in \mathbb{R}$, let $\mathbb{P}_0$ be an observed law on $[d] \times \{0,1\}^2$, and let μ_n be a law on n observed triples. Suppose that:*

- ***(Sampling.)*** *The law μ_n is the n-fold product law generated by $\mathbb{P}_0$, as in Assumption 1.*
- ***(Alphabet size.)*** $2 \le d$.
- ***(Overlap range.)*** $0 < \epsilon < 1/2$.

Then, for every pair $(\mathbf{P}, \mathbf{Q}) \in \Delta_d^2$, let

$$\mathbb{P} = \mathrm{Obs}\{\mathrm{L1Emb}(\mathbf{P}, \mathbf{Q})\},$$

where Obs *is the observed margin from Definition 2 and* L1Emb *is the L_1 embedding from Definition 16. The law $\mathbb{P}$ belongs to $\mathcal{D}_{d,\epsilon}^{\mathrm{obs}}$ in the sense of Definition 3, and, for every $x \in [d]$,*

$$p_x = \frac{P_x + Q_x}{2}.$$

Moreover, whenever $p_x > 0$,

$$\pi_x = \frac{1}{2},$$

and the observed-law optimal-regression value from Definition 4 satisfies

$$\Psi(\mathbb{P}) = \frac{1}{2} + \frac{1}{4} \sum_{x\in[d]} |P_x - Q_x|.$$

Finally, the explicit fixed-n product Markov kernel obtained by applying the single-observation L_1 embedding kernel independently across the n coordinates transforms the fixed paired-sample *law generated by* $(\mathbf{P}, \mathbf{Q})$ *exactly into the product law* $\mathbb{P}^{\otimes n}$. *Consequently, for the fixed-sample minimax squared risk in Definition 5,*

$$\frac{1}{16}\,\mathfrak{R}^{L_1}_{n,d} \le \mathfrak{R}_{n,d,\epsilon},$$

where $\mathfrak{R}^{L_1}_{n,d}$ *is the normalized two-sample* L_1 *minimax mean-squared risk.*

Proposition 3 gives a concrete subexperiment inside the causal-value problem. On this subexperiment, estimating Ψ is exactly as hard as estimating a normalized L_1 distance up to the displayed factor. The Markov-kernel clause makes the comparison fixed-sample and preserves the product sampling structure.

The all-regime lower bound combines this reduction with the moment-matching lower-bound machinery for absolute-value and L_1 functionals in Cai and Low [2011, Lemma 1 and Section 3.1] and Jiao et al. [2018, Theorem 3 and equation (24)].

Theorem 2. [`thm:all-estimator-lower`] (Global minimax lower bound)[1] *Suppose the observed samples satisfy the independent product sampling law in Assumption 1. For every overlap constant* $0 < \epsilon < 1/2$, *there exists a constant* $c_\epsilon > 0$ *such that, for every* $n \ge 1$ *and* $d \ge 2$, *the fixed-sample minimax squared risk* $\mathfrak{R}_{n,d,\epsilon}$ *in Definition 5 satisfies*

$$c_\epsilon \min\left\{1, \frac{d}{n\log(ed)}\right\} \le \mathfrak{R}_{n,d,\epsilon}.$$

The constant c_ϵ depends only on the overlap level. Theorem 2 supplies the information-theoretic scale matched by the Jackson factorial construction: the same expression $d/\{n\log(ed)\}$, truncated at one, governs the worst-case squared risk over the observed model.

Putting the upper and lower inequalities together gives the finite-sample minimax rate. The statement also records the causal-completion equivalence supplied by the observed-margin bridge in Propositions 1 and 2.

Theorem 3. [`thm:matched-minimax-frontier`] (Matched minimax frontier)[2] *There are universal tuning constants* $H_0 > 0$, $0 < \kappa < 1$, *and an integer cutoff* $D_0 > 2$ *for the Jackson factorial estimator such that the following holds. For every fixed overlap constant* $\epsilon \in (0, 1/2)$, *there are real constants* c_ϵ *and* C_ϵ *with*

$$0 < c_\epsilon \le C_\epsilon.$$

For every $n \ge 1$ *and* $d \ge 2$, *with* $L_d = \log(ed)$ *and with* $\mathfrak{R}_{n,d,\epsilon}$ *denoting the fixed-sample observed minimax squared risk in Definition 5,*

$$c_\epsilon \min\left\{1, \frac{d}{nL_d}\right\} \le \mathfrak{R}_{n,d,\epsilon} \le \sup_{\mathbb{P}\in\mathcal{D}^{\mathrm{obs}}_{d,\epsilon}} E_{\mathbb{P}^{\otimes n}}\left[\left(\widehat{V}^{\mathrm{JF}}_{n,d} - \Psi(\mathbb{P})\right)^2\right] \le C_\epsilon \min\left\{1, \frac{d}{nL_d}\right\}.$$

Moreover, the fixed-sample minimax squared risk over the full-data causal completion class $\mathcal{P}_{d,\epsilon}$ *is exactly* $\mathfrak{R}_{n,d,\epsilon}$ *for the same* n, d, *and* ϵ.

[1] **Formalization scope.** This result uses the published conclusion [Cai and Low, 2011] (Lemma 1 and Section 3.1) and [Jiao et al., 2018] (Theorem 3, equation (24), and its Poissonized lower-bound proof). The cited source proof is not formalized here: Lean verifies its use and all remaining steps. If that cited conclusion is false, the portions of this result that depend on it are not certified.

Theorem 3 characterizes the finite-sample minimax order for the observed-law and causal-completion formulations. The upper inequality is attained by the explicit estimator in Algorithm 1; the lower inequality inherits the L_1 approximation and moment-matching ingredients of Cai and Low [2011] and Jiao et al. [2018]. The equality for the causal completion class means that the observed-data risk law is also the risk law for the causal optimal-treatment value.

For asymptotic interpretation, it is useful to name the causal minimax risk directly.

Definition 17. [`synth_4`] (Causal minimax risk) *For $n \geq 1$, $d \geq 2$, and $0 < \epsilon < 1/2$, write $\mathrm{Obs}(P)^{\otimes n}$ for the law of n independent observations each distributed as the observed margin $\mathrm{Obs}(P)$, and define the observed-sample causal minimax squared risk by*

$$\mathfrak{R}^{\mathrm{causal}}_{n,d,\epsilon} := \inf_{\widehat{V}} \sup_{P \in \mathcal{P}_{d,\epsilon}} E_{\mathrm{Obs}(P)^{\otimes n}} \left[\left\{ \widehat{V}(O_1, \ldots, O_n) - V^{\star}(P) \right\}^2 \right],$$

where the infimum is over all measurable maps from $([d] \times \{0,1\}^2)^n$ to $\mathbb{R}$.

The final result translates the finite-sample rate into growth conditions for a sequence of alphabet sizes. Its lower side uses the same published L_1 lower-bound ingredients cited above, and its upper side uses the Jackson factorial estimator with the universal tuning constants from Theorem 3.

Theorem 4. [`thm:consistency-and-parametric-boundaries`] (Consistency and parametric boundaries)[3] *For every overlap constant ϵ with $0 < \epsilon < 1/2$, and for every alphabet-size sequence $d_\bullet = (d_n)_{n \geq 1}$ satisfying $d_n \geq 2$ for all n, the following equivalences hold:*

- ***(Observed consistency.)** The fixed-sample minimax squared risk $\mathfrak{R}_{n,d_n,\epsilon}$ of Definition 5 satisfies*

  $$\mathfrak{R}_{n,d_n,\epsilon} \longrightarrow 0$$

  if and only if

  $$d_n = o\{n \log(en)\}.$$

- ***(Causal consistency.)** The causal minimax squared risk over $\mathcal{P}_{d_n,\epsilon}$ satisfies*

  $$\mathfrak{R}^{\mathrm{causal}}_{n,d_n,\epsilon} \longrightarrow 0$$

  if and only if

  $$d_n = o\{n \log(en)\}.$$

- ***(Observed parametric boundary.)** The observed minimax risk sequence has the parametric n^{-1} rate, meaning that there exist constants c, C with $0 < c \leq C$ such that, for all sufficiently large n,*

  $$\frac{c}{n} \leq \mathfrak{R}_{n,d_n,\epsilon} \leq \frac{C}{n},$$

  if and only if the alphabet-size sequence is uniformly bounded asymptotically, meaning $d_n = O(1)$.

[2] **Formalization scope.** This result uses the published conclusion [Cai and Low, 2011] (Lemma 1 and Section 3.1) and [Jiao et al., 2018] (Theorem 3, equation (24), and its Poissonized lower-bound proof). The cited source proof is not formalized here: Lean verifies its use and all remaining steps. If that cited conclusion is false, the portions of this result that depend on it are not certified.

- ***(Causal parametric boundary.)*** *The causal minimax risk sequence has the parametric n^{-1} rate, meaning that there exist constants c, C with $0 < c \le C$ such that, for all sufficiently large n,*

$$\frac{c}{n} \le \mathfrak{R}^{\mathrm{causal}}_{n,d_n,\epsilon} \le \frac{C}{n},$$

if and only if $d_n = O(1)$.

Theorem 4 gives the operational phase diagram. Uniform consistency corresponds exactly to alphabet growth below $n\log(en)$. The parametric n^{-1} risk scale corresponds exactly to asymptotically bounded alphabet size, for both the observed-law target and the causal optimal-treatment value.

5 Discussion and limitations

The main results give a uniform mean-squared-error scale for the observed-law optimal value over the fixed-overlap model. This section translates that rate into several modeling consequences for categorical adjustment, records the comparison with the predecessor analysis on the same formal class, and isolates the sharp-constant question suggested by the matched upper and lower bounds.

Proposition 4. [`prop:parent-reduction`] (Parent minimax reduction)[4] *There exist tuning constants $H_0 > 0$, $\kappa \in (0,1)$, and an integer cutoff $D_0 > 2$, used by the Jackson factorial estimator in Algorithm 1, such that:*

- ***(Matched frontier.)*** *For every overlap parameter $\epsilon \in (0, 1/2)$, there are constants c and C with $0 < c \le C$ such that, for every $n \ge 1$ and $d \ge 2$,*

$$c\min\left\{1, \frac{\sqrt{d}}{n} + \frac{d}{n\log(en)}\right\} \le \mathfrak{R}_{n,d,\epsilon} \le C\frac{d}{n}.$$

- ***(Bounded alphabets.)*** *For every n, every $2 \le d < D_0$, every overlap parameter ϵ, and every observed sample, the Jackson factorial estimator in Algorithm 1 equals the empirical-ratio estimator in Definition 10:*

$$\widehat{V}^{\mathrm{JF}}_{n,d} = \widehat{V}^{\mathrm{ER}}_{n,d}.$$

- ***(Bounded-alphabet sequences.)*** *For every overlap parameter $\epsilon \in (0, 1/2)$ and every sequence d_n with $d_n \ge 2$ for all n and $d_n \le d_{\max}$ for some finite $d_{\max}$, there are constants c and C with $0 < c \le C$ such that, for all sufficiently large n,*

$$\frac{c}{n} \le \mathfrak{R}_{n,d_n,\epsilon} \le \frac{C}{n}.$$

[3] **Formalization scope.** This result uses the published conclusion [Cai and Low, 2011] (Lemma 1 and Section 3.1) and [Jiao et al., 2018] (Theorem 3, equation (24), and its Poissonized lower-bound proof). The cited source proof is not formalized here: Lean verifies its use and all remaining steps. If that cited conclusion is false, the portions of this result that depend on it are not certified.

[4] **Formalization scope.** This result uses the published conclusion [Cai and Low, 2011] (Lemma 1 and Section 3.1) and [Jiao et al., 2018] (Theorem 3, equation (24), and its Poissonized lower-bound proof). The cited source proof is not formalized here: Lean verifies its use and all remaining steps. If that cited conclusion is false, the portions of this result that depend on it are not certified.

Proposition 4 places the matched-rate analysis on the same observed model class and target used by the earlier bracket. Its first clause is labelled *Matched frontier* because it is a consequence of the matched frontier, and not because the two displayed bounds meet: the lower expression $\sqrt{d}/n+d/\{n\log(en)\}$ and the upper expression Cd/n are the predecessor's nonmatching bracket, which Theorem 3 implies and strictly sharpens. The clause is recorded here so that the earlier bracket can be read off the present theorem on the same formal class; the sharp comparison is Theorem 3 alone. The second clause explains how the constructed estimator is anchored at small alphabets, and the third clause records the fixed-dimensional parametric consequence in sequence form. The external ingredients are the classical moment-matching approximation to $|t|$ in Cai and Low [2011, Lemma 1 and Section 3.1] and the two-sample L_1 lower-bound mechanism in Jiao et al. [2018, Theorem 3, equation (24)].

The statistical message is that, once fixed overlap is imposed, alphabet growth is what determines the nonparametric difficulty. Exact treatment-effect ties create the local nonsmoothness of the cellwise maximum, and the Jackson factorial construction in Algorithm 1 smooths that nonsmoothness at the logarithmic degree dictated by the alphabet scale. Unequal propensities enter through the overlap-stabilized representation in Definition 8: fixed overlap keeps the observed arm masses comparable enough for the same rate calculation, while the constants may depend on the overlap parameter.

Null cells and boundary outcome means are handled inside the model and estimator definitions themselves. The observed model class in Definition 3 permits zero covariate-cell mass, and the empirical-ratio estimator in Definition 10 fixes the zero-over-zero convention used by the bounded-alphabet branch. At the boundary of the binary outcome space, the global extension in Definition 8 keeps the cell contribution defined on the nonnegative orthant, so the same risk statement covers outcome means equal to zero or one.

For bounded alphabets, Proposition 4 gives the usual n^{-1} minimax scale after a finite burn-in along the sequence. This confirms that the logarithmic approximation gain is a growing-alphabet phenomenon: when the covariate alphabet is uniformly bounded, the optimal-value problem has the same squared-risk order as an ordinary finite-dimensional regular estimation problem, with constants depending on the alphabet bound and overlap.

5.1 Limitations and future work

The established domain is deliberately narrow. Every result in this paper is proved for a binary treatment, a binary outcome, a finite covariate alphabet, and a fixed overlap constant $\epsilon \in (0,1/2)$, with the comparison constants c_ϵ and C_ϵ depending on ϵ; the target is a scalar and the loss is squared error; and the number of treatment arms is two and does not grow. Four directions therefore sit outside the present theorems. Overlap that vanishes with the sample size changes the effective per-cell information and is not covered by constants indexed by a fixed ϵ. A growing number of treatment arms replaces the two-arm maximum by a maximum over a growing set, which changes the approximation problem the Jackson degree is calibrated to. Continuous or unbounded outcomes, and continuous covariates, leave the finite four-atom cell representation on which the estimator and the embedding are built. Finally, the paper gives a point estimator and a squared-risk frontier; uniform confidence procedures for the optimal value under the same growing-alphabet asymptotics are a separate question, and the nonregularity at treatment-effect ties is what makes it a hard one.

The matched comparison constants leave open an exact efficiency calculation for nonsaturated growing alphabets. The natural normalization is determined by the $d/\{n\log(ed)\}$ term in Theorem 3; an exact constant would identify the efficiency price of estimating the cellwise maximum under unknown observational treatment imbalance.

Definition 18. [`def:constant-handle`] (Constant route $\mathscr{H}_\epsilon^{\mathrm{const}}$) *The route $\mathscr{H}_\epsilon^{\mathrm{const}}$ rescales the pilot-local four-cell rectangles by their Poisson noise geometry, optimizes the tensor-polynomial bias–variance functional for $\bar{f}_\epsilon$, and dualizes the approximation constraint into moment-matched observed laws in $\mathcal{D}_{d,\epsilon}^{\mathrm{obs}}$ for the target $\Psi(\mathbb{P})$.*

Definition 18 names the analytic route suggested by the upper-bound construction and the lower-bound duality. It combines the local Poisson geometry used by the Jackson factorial estimator with the approximation problem for $\bar{f}_\epsilon$, the globally defined cell contribution from Definition 8. This route is the candidate mechanism for moving from comparison constants to a sharp overlap-dependent constant.

Remark 1. [`oeq:sharp-overlap-constant`] (Sharp overlap constant) *A natural next question is whether, along each nonsaturated sequence with $d \to \infty$ and*

$$\frac{d}{n\log(ed)} \to 0,$$

the normalized risk

$$\frac{n\log(ed)}{d}\mathfrak{R}_{n,d,\epsilon}$$

converges to a finite positive limit, and whether an estimator attaining that limit can be derived through $\mathscr{H}_\epsilon^{\mathrm{const}}$.

Remark 1 states the resulting open question. The nonsaturated condition keeps the risk on the vanishing side of the consistency boundary in Theorem 4, and the normalization matches the leading growing-alphabet term in Theorem 3. A positive answer would refine the present minimax rate into an exact asymptotic constant for fixed overlap.

Appendices

A Analytic ingredients for the upper bound

The upper bound in Theorem 1 rests on two local analytic devices. The first is a Jackson polynomial certificate for the globally defined cell contribution $\bar{f}_\epsilon$ from Definition 8, localized to the pilot rectangle Q_x constructed in Algorithm 1. The second is a centered Poisson factorial lift that turns the local polynomial into a cell statistic whose bias and variance remain controlled at the cell mass scale.

We begin with the approximation component. The construction uses the four outcome-arm coordinate set $\mathcal{J}$ and smooths a bounded continuous function on the normalized cube by the tensor product of the order-four Jackson kernel in Definition 12. The following coefficient envelope records the degree and one-norm control needed when the smoothed polynomial is later evaluated by factorial moments.

Lemma 1. [`lem:tensor-jackson-coefficient-envelope`] (Tensor Jackson coefficient envelope) *With the Jackson kernel notation of Definition 12, assume:*

- ***(Order.)*** *$K \geq 1$ is an integer.*

- ***(Function.)*** *f is continuous on the normalized cube $[-1,1]^{\mathcal{J}}$, where $\mathcal{J}$ is the four arm-outcome coordinate set.*

- ***(Envelope.)*** *$B \geq 0$ and $|f(z)| \leq B$ for every $z \in [-1,1]^{\mathcal{J}}$.*

Then there exists a polynomial $\mathcal{R}$ in the four normalized coordinates such that, for every angle vector $\varphi : \mathcal{J} \to \mathbb{R}$,

$$\mathcal{R}(\cos\varphi) = \int_{[-\pi,\pi]^{\mathcal{J}}} f\{\cos(\varphi - \tau)\} \prod_{\jmath \in \mathcal{J}} J_K(\tau_\jmath)\, d\tau.$$

Here cosine is applied coordinatewise. Moreover, every multi-index $\beta \in \operatorname{supp}(\mathcal{R})$ satisfies

$$\beta_\omega \leq 2(K-1) \qquad \textit{for every } \omega \in \mathcal{J},$$

the total degree of $\mathcal{R}$ is at most $8(K-1)$, and its coefficient one-norm satisfies

$$\sum_{\beta \in \operatorname{supp}(\mathcal{R})} |\mathcal{R}_\beta| \leq 2^{40K+20} B.$$

Proof of Lemma 1. Write

$$\mathbb{T}_{\mathcal{J}} := [-\pi,\pi]^{\mathcal{J}}, \qquad N_K := 2(K-1),$$

and define the tensor convolution

$$(\mathcal{J}_K f)(\varphi) := \int_{\mathbb{T}_{\mathcal{J}}} f\{\cos(\varphi - \tau)\} \prod_{\jmath \in \mathcal{J}} J_K(\tau_\jmath)\, d\tau.$$

For the positive order K in the lemma, write $J_K = c_K G_K$, where

$$G_K(t) = \begin{cases} \left\{ \dfrac{\sin(Kt/2)}{\sin(t/2)} \right\}^4, & \sin(t/2) \neq 0, \\ K^4, & \sin(t/2) = 0, \end{cases}$$

with the removable values filled continuously, and set

$$c_K := \left(\int_{-\pi}^{\pi} G_K(t)\, dt \right)^{-1}.$$

Then $G_K(t) \geq 0$ for every t, and $G_K(0) = K^4 > 0$, so $\int_{-\pi}^{\pi} G_K(t)\, dt > 0$ and $c_K > 0$. Thus $J_K \geq 0$ and

$$\int_{-\pi}^{\pi} J_K(t)\, dt = c_K \int_{-\pi}^{\pi} G_K(t)\, dt = 1.$$

Fubini's theorem for the nonnegative continuous tensor product then gives

$$\prod_{\jmath \in \mathcal{J}} J_K(\tau_\jmath) \geq 0, \qquad \int_{\mathbb{T}_{\mathcal{J}}} \prod_{\jmath \in \mathcal{J}} J_K(\tau_\jmath)\, d\tau = 1.$$

Continuity of f on the normalized cube and continuity of the coordinatewise cosine map give the integrability needed for the displayed convolutions and for the linearity identities below.

1. First construct the polynomial representation. The Jackson frequency bound follows from the Chebyshev form of the raw kernel. If U_{K-1} denotes the Chebyshev polynomial of the second kind, then

$$G_K(s) = \{U_{K-1}(\cos(s/2))\}^4.$$

Since $U_{K-1}(-\varrho) = (-1)^{K-1}U_{K-1}(\varrho)$, the fourth power is an even polynomial in ϱ. Hence there is a polynomial Q_K^{even} of degree at most $2(K-1) = N_K$ such that

$$\{U_{K-1}(\varrho)\}^4 = Q_K^{\mathrm{even}}(\varrho^2).$$

Using $\cos^2(s/2) = (1+\cos s)/2$, this gives

$$G_K(s) = Q_K^{\mathrm{even}}\{(1+\cos s)/2\}.$$

Multiplication by the constant c_K preserves the frequency bound, and a polynomial of degree at most N_K in $\cos s$ is a trigonometric polynomial with frequencies at most N_K. Thus there are real coefficient sequences $\chi_{K,r}^{\cos}$ and $\chi_{K,r}^{\sin}$, $0 \le r \le N_K$, such that

$$J_K(s) = \sum_{r=0}^{N_K}\{\chi_{K,r}^{\cos}\cos(rs) + \chi_{K,r}^{\sin}\sin(rs)\}.$$

Fix all coordinates except $\omega_0 \in \mathcal{J}$. For an angle vector $\eta : \mathcal{J} \to \mathbb{R}$ and $t \in \mathbb{R}$, write

$$(\eta_{\omega_0} \leftarrow t)_\omega = \begin{cases} t, & \omega = \omega_0, \\ \eta_\omega, & \omega \neq \omega_0. \end{cases}$$

Set, for $\psi \in \mathbb{T}_{\mathcal{J}}$,

$$W_{\omega_0,\eta}(\psi) := f(\cos\psi) \prod_{\omega\in\mathcal{J}\setminus\{\omega_0\}} J_K(\eta_\omega - \psi_\omega).$$

The period-box translation $\psi = (\eta_{\omega_0} \leftarrow t) - \tau$, with the box understood modulo the common 2π-period in each coordinate, gives

$$(\mathcal{J}_K f)(\eta_{\omega_0} \leftarrow t) = \int_{\mathbb{T}_{\mathcal{J}}} f(\cos\psi) \prod_{\omega\in\mathcal{J}} J_K\{(\eta_{\omega_0} \leftarrow t)_\omega - \psi_\omega\}\, d\psi = \int_{\mathbb{T}_{\mathcal{J}}} W_{\omega_0,\eta}(\psi) J_K(t-\psi_{\omega_0})\, d\psi.$$

The translation is valid because the displayed integrand is continuous and is 2π-periodic in each coordinate; the periodicity follows from the cosine term and the trigonometric-polynomial representation of J_K. Define

$$\Xi_{\omega_0,\eta,r}^{\cos} := \int_{\mathbb{T}_{\mathcal{J}}} W_{\omega_0,\eta}(\psi)\{\chi_{K,r}^{\cos}\cos(r\psi_{\omega_0}) - \chi_{K,r}^{\sin}\sin(r\psi_{\omega_0})\}\, d\psi,$$

$$\Xi_{\omega_0,\eta,r}^{\sin} := \int_{\mathbb{T}_{\mathcal{J}}} W_{\omega_0,\eta}(\psi)\{\chi_{K,r}^{\cos}\sin(r\psi_{\omega_0}) + \chi_{K,r}^{\sin}\cos(r\psi_{\omega_0})\}\, d\psi.$$

Substituting the finite expansion of $J_K(t-\psi_{\omega_0})$, using the angle-subtraction identities for sine and cosine, and integrating the finite sum term by term gives

$$(\mathcal{J}_K f)(\eta_{\omega_0} \leftarrow t) = \sum_{r=0}^{N_K}\{\Xi_{\omega_0,\eta,r}^{\cos}\cos(rt) + \Xi_{\omega_0,\eta,r}^{\sin}\sin(rt)\}.$$

Thus each one-coordinate section is a trigonometric polynomial of frequency at most N_K. The same section is even in that coordinate: reflecting the integration variable $\tau_{\omega_0} \mapsto -\tau_{\omega_0}$ preserves $\mathbb{T}_{\mathcal{J}}$, the kernel is even, and cosine is even. Therefore the sine part cancels. Interpolating the resulting even trigonometric polynomial in one cosine coordinate at the nodes

$$\xi_j = \frac{j}{N_K+1}, \qquad \zeta_j = \arccos(\xi_j), \qquad j = 0, \dots, N_K,$$

and repeating this coordinate by coordinate gives a polynomial $\mathcal{R}$ in the four variables such that

$$\mathcal{R}(\cos\varphi) = (\mathcal{J}_K f)(\varphi) \quad \text{for every } \varphi : \mathcal{J} \to \mathbb{R}.$$

The interpolation basis has degree at most N_K in the coordinate being added, while the induction hypothesis controls the other coordinates. Hence, for every $\beta \in \operatorname{supp}(\mathcal{R})$,

$$\beta_\omega \le N_K = 2(K-1) \quad \text{for every } \omega \in \mathcal{J},$$

and

$$\deg(\mathcal{R}) \le 4N_K = 8(K-1).$$

2. The convolution is bounded by the same envelope B. Since $|f(z)| \le B$ on the normalized cube,

$$B - f(z) \ge 0, \qquad B + f(z) \ge 0 \quad (z \in [-1,1]^{\mathcal{J}}).$$

Kernel nonnegativity gives

$$(\mathcal{J}_K(B-f))(\varphi) \ge 0, \qquad (\mathcal{J}_K(B+f))(\varphi) \ge 0.$$

By linearity of the integral and the unit-mass identity,

$$(\mathcal{J}_K(B-f))(\varphi) = B - (\mathcal{J}_K f)(\varphi), \qquad (\mathcal{J}_K(B+f))(\varphi) = B + (\mathcal{J}_K f)(\varphi).$$

Therefore

$$-B \le (\mathcal{J}_K f)(\varphi) \le B, \qquad |(\mathcal{J}_K f)(\varphi)| \le B \quad (\varphi : \mathcal{J} \to \mathbb{R}).$$

3. This angle-space bound transfers to the whole normalized cube. For any $z \in [-1,1]^{\mathcal{J}}$, take

$$\varphi_z(\omega) := \arccos(z_\omega), \qquad \omega \in \mathcal{J},$$

so that $\cos(\varphi_z)_\omega = z_\omega$. The representation from the first step and the bound from the second step give

$$|\mathcal{R}(z)| = |\mathcal{R}(\cos\varphi_z)| = |(\mathcal{J}_K f)(\varphi_z)| \le B.$$

4. It remains to bound the coefficient one-norm. The coefficient estimate used here is the following. Let $\mathcal{S}$ be a polynomial in δ_{dim} normalized variables, each coordinate degree at most N_K, and suppose

$$|\mathcal{S}(y)| \le B_{\mathrm{env}} \qquad (y \in [-1,1]^{\delta_{\mathrm{dim}}}), \qquad B_{\mathrm{env}} \ge 0.$$

Then

$$\sum_{\rho \in \operatorname{supp}(\mathcal{S})} |\mathcal{S}_\rho| \le \{2(N_K+1)^2 3^{N_K}\}^{\delta_{\mathrm{dim}}} B_{\mathrm{env}}.$$

For one variable, if s has degree at most N_K and $|s(x)| \le C$ on $[-1,1]$, then $s(\cos t)$ is an even trigonometric polynomial,

$$s(\cos t) = \sum_{r=0}^{N_K} a_r \cos(rt).$$

Orthogonality on $[-\pi, \pi]$ gives

$$|a_r| \le 2C \qquad (0 \le r \le N_K).$$

Since $\cos(rt) = T_r(\cos t)$ and the Chebyshev coefficient one-norm satisfies

$$\|T_r\|_1 \le 3^r,$$

the polynomial identity $s(x) = \sum_{r=0}^{N_K} a_r T_r(x)$ yields

$$\|s\|_1 \le \sum_{r=0}^{N_K} |a_r| \, \|T_r\|_1 \le 2(N_K+1)3^{N_K} C.$$

For the multivariate estimate, write

$$\mathcal{S}(x_0, y) = \sum_{r=0}^{N_K} s_r(y) x_0^r.$$

For each fixed $y \in [-1,1]^{\delta_{\mathrm{dim}}}$, the preceding one-dimensional bound gives

$$|s_r(y)| \le 2(N_K+1)3^{N_K} B_{\mathrm{env}}.$$

Applying the induction hypothesis to each coefficient polynomial s_r and summing over $r = 0, \ldots, N_K$ gives

$$\|\mathcal{S}\|_1 = \sum_{r=0}^{N_K} \|s_r\|_1 \le (N_K+1)\{2(N_K+1)^2 3^{N_K}\}^{\delta_{\mathrm{dim}}} \{2(N_K+1)3^{N_K} B_{\mathrm{env}}\},$$

which is exactly

$$\|\mathcal{S}\|_1 \le \{2(N_K+1)^2 3^{N_K}\}^{\delta_{\mathrm{dim}}+1} B_{\mathrm{env}}.$$

The case $\delta_{\mathrm{dim}} = 0$ is the constant-polynomial case and follows directly from the displayed uniform bound.

5. Apply this coefficient estimate to $\mathcal{S} = \mathcal{R}$, $\delta_{\mathrm{dim}} = 4$, and $B_{\mathrm{env}} = B$. With $N_K = 2(K-1)$,

$$\sum_{\beta \in \mathrm{supp}(\mathcal{R})} |\mathcal{R}_\beta| \le \{2(N_K+1)^2 3^{N_K}\}^4 B.$$

The elementary bound

$$N_K + 1 \le 2^{N_K+1}$$

and $3^{N_K} \le 4^{N_K}$ imply

$$2(N_K+1)^2 3^{N_K} \le 2\{2^{N_K+1}\}^2 4^{N_K} = 2^{4N_K+3}.$$

Consequently

$$\{2(N_K+1)^2 3^{N_K}\}^4 \le 2^{4(4N_K+3)}.$$

Since $N_K = 2(K-1)$ and $K \ge 1$,

$$4(4N_K+3) = 32(K-1)+12 \le 40K+20.$$

Because the base 2 is at least 1 and $B \ge 0$,

$$\sum_{\beta \in \operatorname{supp}(\mathcal{R})} |\mathcal{R}_\beta| \le 2^{40K+20} B.$$

Together with the representation, coordinate-support bound, and total-degree bound proved above, this is the asserted polynomial $\mathcal{R}$.

□

The envelope in Lemma 1 is deliberately expressed in normalized coordinates. After an affine change of variables from the pilot rectangle to $[-1,1]^{\mathcal{J}}$, it gives a physical polynomial $P_{x,K}$ while preserving a centered expansion around b_x, the rectangle center from Definition 12.

Lemma 2. [`lem:jackson-tensor-extraction`] (Jackson tensor extraction) *Using the overlap-parameter notation of Assumption 4, let $\epsilon \in \mathbb{R}$, let K be an integer, and let $Q_x = \prod_{\jmath \in \mathcal{J}} [\ell_{\jmath,x}, u_{\jmath,x}]$ be a four-coordinate rectangle with*

$$b_{\jmath,x} = \frac{\ell_{\jmath,x} + u_{\jmath,x}}{2}, \qquad r_{\jmath,x} = \frac{u_{\jmath,x} - \ell_{\jmath,x}}{2}.$$

Assume:

- ***(Degree.)*** *$K \ge 2$.*
- ***(Lower endpoints.)*** *$\ell_{\jmath,x} \ge 0$ for every $\jmath \in \mathcal{J}$.*
- ***(Overlap parameter.)*** *$\epsilon > 0$.*
- ***(Radii.)*** *$r_{\jmath,x} > 0$ for every $\jmath \in \mathcal{J}$.*

Then there exist a polynomial $P_{x,K}$ in the four physical coordinates and a polynomial $\mathcal{R}_{x,K}$ in the four normalized coordinates such that:

- ***(Physical degree.)*** *Every coordinate degree of $P_{x,K}$ is at most $2(K-1)$.*
- ***(Normalized degree.)*** *Every $\beta \in \operatorname{supp}(\mathcal{R}_{x,K})$ satisfies $\beta_\omega \le 2(K-1)$ for every coordinate $\omega \in \mathcal{J}$.*
- ***(Centered evaluation.)*** *For every four-coordinate vector v,*

$$P_{x,K}(v) - \bar{f}_\epsilon(b_x) = \mathcal{R}_{x,K}\left(\left\{\frac{v_\jmath - b_{\jmath,x}}{r_{\jmath,x}}\right\}_{\jmath \in \mathcal{J}}\right).$$

Here $\bar{f}_\epsilon$ is the globally defined cellwise maximum contribution from Definition 8.

- ***(Jackson convolution.)*** *For every angle vector $\vartheta : \mathcal{J} \to \mathbb{R}$,*

$$P_{x,K}\left(\{b_{\jmath,x} + r_{\jmath,x}\cos(\vartheta_\jmath)\}_{\jmath\in\mathcal{J}}\right)$$

  equals the order-K tensor Jackson convolution from Definition 12 of $\bar{f}_\epsilon$ precomposed with the affine parametrization of Q_x, evaluated at ϑ.

- ***(Coefficient envelope.)***

$$\sum_{\beta\in\operatorname{supp}(\mathcal{R}_{x,K})} |(\mathcal{R}_{x,K})_\beta| \le 2^{40K+20}(1+\epsilon^{-1})\sum_{\jmath\in\mathcal{J}} r_{\jmath,x}.$$

Proof of Lemma 2. 1. Define

$$\mathrm{Box}_x = \{y \in \mathbb{R}^{\mathcal{J}} : |y_\jmath - b_{\jmath,x}| \le r_{\jmath,x} \text{ for every } \jmath \in \mathcal{J}\}, \qquad \mathrm{Cube}_{\mathcal{J}} = [-1,1]^{\mathcal{J}},$$

and

$$\mathrm{aff}_x(\zeta)_\jmath = b_{\jmath,x} + r_{\jmath,x}\zeta_\jmath, \qquad \zeta \in \mathbb{R}^{\mathcal{J}}.$$

If $y \in \mathrm{Box}_x$, then

$$y_\jmath \ge b_{\jmath,x} - r_{\jmath,x} = \frac{\ell_{\jmath,x} + u_{\jmath,x}}{2} - \frac{u_{\jmath,x} - \ell_{\jmath,x}}{2} = \ell_{\jmath,x} \ge 0 \qquad (\jmath \in \mathcal{J}),$$

and $b_x \in \mathrm{Box}_x$, since each $r_{\jmath,x} > 0$.

For nonnegative four-coordinate vectors u, w, put

$$\delta_{\mathcal{J}}(u,w) = \sum_{\jmath\in\mathcal{J}} |u_\jmath - w_\jmath|.$$

For each arm $a \in \{0,1\}$, Definition 8 gives

$$g_{a,\epsilon}(u) = \begin{cases} 0, & s(u) = 0, \\[2ex] \dfrac{s(u)u_{a1}}{\max\{s_a(u), \epsilon s(u)\}}, & s(u) > 0. \end{cases}$$

Moreover

$$0 \le g_{a,\epsilon}(u) \le s(u),$$

because $u_{a1} \le s_a(u) \le \max\{s_a(u), \epsilon s(u)\}$ when $s(u) > 0$, and the zero-mass value is 0. If $s(u) = 0$, nonnegativity gives $u = 0$, hence

$$|g_{a,\epsilon}(u) - g_{a,\epsilon}(w)| = g_{a,\epsilon}(w) \le s(w) = \delta_{\mathcal{J}}(u,w) \le (1+\epsilon^{-1})\delta_{\mathcal{J}}(u,w).$$

The case $s(w) = 0$ is symmetric.

Assume now that $s(u) > 0$ and $s(w) > 0$. The elementary bounds

$$|s(u) - s(w)| \le \delta_{\mathcal{J}}(u,w), \qquad |u_{a1} - w_{a1}| \le \delta_{\mathcal{J}}(u,w),$$

and

$$|s_a(u) - s_a(w)| \le |u_{a0} - w_{a0}| + |u_{a1} - w_{a1}| \le \delta_{\mathcal{J}}(u,w)$$

will be used throughout. If

$$s_a(u) \le \epsilon s(u), \qquad s_a(w) \le \epsilon s(w),$$

then

$$g_{a,\epsilon}(u) = \frac{u_{a1}}{\epsilon}, \qquad g_{a,\epsilon}(w) = \frac{w_{a1}}{\epsilon},$$

so

$$|g_{a,\epsilon}(u) - g_{a,\epsilon}(w)| \le \epsilon^{-1}\delta_{\mathcal{J}}(u,w) \le (1+\epsilon^{-1})\delta_{\mathcal{J}}(u,w).$$

If

$$\epsilon s(u) \le s_a(u), \qquad \epsilon s(w) \le s_a(w),$$

then

$$g_{a,\epsilon}(u) = s(u)\frac{u_{a1}}{s_a(u)}, \qquad g_{a,\epsilon}(w) = s(w)\frac{w_{a1}}{s_a(w)}.$$

Since $0 \le w_{a1}/s_a(w) \le 1$ and $s(u)/s_a(u) \le \epsilon^{-1}$, while

$$\left|\frac{u_{a1}}{s_a(u)} - \frac{w_{a1}}{s_a(w)}\right| \le \frac{|u_{a1} - w_{a1}| + |u_{a0} - w_{a0}|}{s_a(u)},$$

we obtain

$$\begin{aligned}|g_{a,\epsilon}(u) - g_{a,\epsilon}(w)| &\le |s(u) - s(w)| + s(u)\left|\frac{u_{a1}}{s_a(u)} - \frac{w_{a1}}{s_a(w)}\right| \\ &\le \delta_{\mathcal{J}}(u,w) + \epsilon^{-1}\delta_{\mathcal{J}}(u,w) = (1+\epsilon^{-1})\delta_{\mathcal{J}}(u,w).\end{aligned}$$

If

$$s_a(u) \le \epsilon s(u), \qquad \epsilon s(w) \le s_a(w),$$

then

$$g_{a,\epsilon}(u) - g_{a,\epsilon}(w) = \frac{u_{a1} - w_{a1}}{\epsilon} + \frac{w_{a1}}{s_a(w)}\frac{s_a(w) - \epsilon s(w)}{\epsilon}.$$

The second summand is nonnegative, giving

$$g_{a,\epsilon}(u) - g_{a,\epsilon}(w) \ge -\epsilon^{-1}\delta_{\mathcal{J}}(u,w).$$

For the upper bound, $0 \le w_{a1}/s_a(w) \le 1$ and

$$0 \le s_a(w) - \epsilon s(w) \le \{s_a(w) - s_a(u)\} - \epsilon\{s(w) - s(u)\},$$

so

$$\begin{aligned}g_{a,\epsilon}(u) - g_{a,\epsilon}(w) &\le \frac{u_{a1} - w_{a1} + s_a(w) - s_a(u) - \epsilon\{s(w) - s(u)\}}{\epsilon} \\ &= \epsilon^{-1}(w_{a0} - u_{a0}) - \{s(w) - s(u)\} \\ &\le (1+\epsilon^{-1})\delta_{\mathcal{J}}(u,w).\end{aligned}$$

The remaining mixed case,

$$\epsilon s(u) \le s_a(u), \qquad s_a(w) \le \epsilon s(w),$$

is analogous. Here

$$g_{a,\epsilon}(u) - g_{a,\epsilon}(w) = \frac{u_{a1} - w_{a1}}{\epsilon} - \frac{u_{a1}}{s_a(u)} \frac{s_a(u) - \epsilon s(u)}{\epsilon}.$$

The subtracted summand is nonnegative, hence

$$g_{a,\epsilon}(u) - g_{a,\epsilon}(w) \le \epsilon^{-1} \delta_{\mathcal{J}}(u, w).$$

Also $0 \le u_{a1}/s_a(u) \le 1$ and

$$0 \le s_a(u) - \epsilon s(u) \le \{s_a(u) - s_a(w)\} - \epsilon\{s(u) - s(w)\},$$

which yields

$$\begin{aligned}
g_{a,\epsilon}(u) - g_{a,\epsilon}(w) &\ge \frac{u_{a1} - w_{a1} - s_a(u) + s_a(w) + \epsilon\{s(u) - s(w)\}}{\epsilon} \\
&= \epsilon^{-1}(w_{a0} - u_{a0}) + \{s(u) - s(w)\} \\
&\ge -(1 + \epsilon^{-1}) \delta_{\mathcal{J}}(u, w).
\end{aligned}$$

Thus, for both arms,

$$|g_{a,\epsilon}(u) - g_{a,\epsilon}(w)| \le (1 + \epsilon^{-1}) \delta_{\mathcal{J}}(u, w).$$

Since

$$|\max\{\rho_0, \rho_1\} - \max\{\sigma_0, \sigma_1\}| \le \max_{a \in \{0,1\}} |\rho_a - \sigma_a|,$$

Definition 8 gives

$$|\bar{f}_\epsilon(u) - \bar{f}_\epsilon(w)| \le (1 + \epsilon^{-1}) \delta_{\mathcal{J}}(u, w).$$

Consequently $\bar{f}_\epsilon$ is continuous on Box_x, and for every $y \in \mathrm{Box}_x$,

$$|\bar{f}_\epsilon(y) - \bar{f}_\epsilon(b_x)| \le (1 + \epsilon^{-1}) \sum_{\jmath \in \mathcal{J}} |y_\jmath - b_{\jmath,x}| \le (1 + \epsilon^{-1}) \sum_{\jmath \in \mathcal{J}} r_{\jmath,x}.$$

Set

$$S_x = (1 + \epsilon^{-1}) \sum_{\jmath \in \mathcal{J}} r_{\jmath,x}.$$

2. Define the centered pullback on the normalized cube by

$$\chi_x(\zeta) = \bar{f}_\epsilon(\mathrm{aff}_x(\zeta)) - \bar{f}_\epsilon(b_x), \qquad \zeta \in \mathrm{Cube}_{\mathcal{J}}.$$

The affine map sends $\mathrm{Cube}_{\mathcal{J}}$ into Box_x, so χ_x is continuous and

$$|\chi_x(\zeta)| \le S_x \qquad (\zeta \in \mathrm{Cube}_{\mathcal{J}}).$$

Since $K \ge 2$, we have $K \ge 1$. Applying Lemma 1 to χ_x with envelope $B = S_x$ gives a polynomial $\mathcal{R}_{x,K}$ in the normalized coordinates such that

$$\mathcal{R}_{x,K}(\cos\vartheta) = \int_{[-\pi,\pi]^{\mathcal{J}}} \chi_x\{\cos(\vartheta - \tau)\} \prod_{\jmath \in \mathcal{J}} J_K(\tau_\jmath)\, d\tau \qquad (\vartheta : \mathcal{J} \to \mathbb{R}),$$

$$\beta_\omega \le 2(K-1) \quad \text{for every } \beta \in \mathrm{supp}(\mathcal{R}_{x,K}) \text{ and } \omega \in \mathcal{J},$$

and

$$\sum_{\beta \in \mathrm{supp}(\mathcal{R}_{x,K})} |(\mathcal{R}_{x,K})_\beta| \le 2^{40K+20} S_x.$$

3. Substitute normalized coordinates into physical coordinates. Since every $r_{\jmath,x} > 0$, there is a polynomial $P^0_{x,K}$ in the four physical coordinates satisfying

$$P^0_{x,K}(v) = \mathcal{R}_{x,K}\left(\left\{\frac{v_\jmath - b_{\jmath,x}}{r_{\jmath,x}}\right\}_{\jmath\in\mathcal{J}}\right) \qquad (v \in \mathbb{R}^{\mathcal{J}}),$$

and every coordinate degree of $P^0_{x,K}$ is at most $2(K-1)$. Define

$$P_{x,K}(v) = P^0_{x,K}(v) + \bar{f}_\epsilon(b_x).$$

Adding a constant preserves the coordinate-degree bound, and therefore

$$P_{x,K}(v) - \bar{f}_\epsilon(b_x) = \mathcal{R}_{x,K}\left(\left\{\frac{v_\jmath - b_{\jmath,x}}{r_{\jmath,x}}\right\}_{\jmath\in\mathcal{J}}\right) \qquad (v \in \mathbb{R}^{\mathcal{J}}).$$

4. For an angle vector $\vartheta : \mathcal{J} \to \mathbb{R}$,

$$\left\{\frac{b_{\jmath,x} + r_{\jmath,x}\cos(\vartheta_\jmath) - b_{\jmath,x}}{r_{\jmath,x}}\right\}_{\jmath\in\mathcal{J}} = \{\cos(\vartheta_\jmath)\}_{\jmath\in\mathcal{J}}.$$

Thus

$$P_{x,K}(b_x + r_x \odot \cos\vartheta) = \mathcal{R}_{x,K}(\cos\vartheta) + \bar{f}_\epsilon(b_x).$$

Using the convolution identity for $\mathcal{R}_{x,K}$,

$$\begin{aligned} P_{x,K}(b_x + r_x \odot \cos\vartheta) = \int_{[-\pi,\pi]^{\mathcal{J}}} &\{\bar{f}_\epsilon(b_x + r_x \odot \cos(\vartheta - \tau)) - \bar{f}_\epsilon(b_x)\} \prod_{\jmath\in\mathcal{J}} J_K(\tau_\jmath)\, d\tau \\ &+ \bar{f}_\epsilon(b_x). \end{aligned}$$

The integrands are continuous on the compact period box, and the tensor Jackson kernel has unit mass,

$$\int_{[-\pi,\pi]^{\mathcal{J}}} \prod_{\jmath\in\mathcal{J}} J_K(\tau_\jmath)\, d\tau = 1.$$

Therefore the constant centered term integrates exactly to the value added back:

$$\begin{aligned} &\int_{[-\pi,\pi]^{\mathcal{J}}} \{\bar{f}_\epsilon(b_x + r_x \odot \cos(\vartheta - \tau)) - \bar{f}_\epsilon(b_x)\} \prod_{\jmath\in\mathcal{J}} J_K(\tau_\jmath)\, d\tau + \bar{f}_\epsilon(b_x) \\ &\quad = \int_{[-\pi,\pi]^{\mathcal{J}}} \bar{f}_\epsilon(b_x + r_x \odot \cos(\vartheta - \tau)) \prod_{\jmath\in\mathcal{J}} J_K(\tau_\jmath)\, d\tau - \bar{f}_\epsilon(b_x) \int_{[-\pi,\pi]^{\mathcal{J}}} \prod_{\jmath\in\mathcal{J}} J_K(\tau_\jmath)\, d\tau + \bar{f}_\epsilon(b_x) \\ &\quad = \int_{[-\pi,\pi]^{\mathcal{J}}} \bar{f}_\epsilon(b_x + r_x \odot \cos(\vartheta - \tau)) \prod_{\jmath\in\mathcal{J}} J_K(\tau_\jmath)\, d\tau. \end{aligned}$$

Thus

$$P_{x,K}(b_x + r_x \odot \cos\vartheta) = \int_{[-\pi,\pi]^{\mathcal{J}}} \bar{f}_\epsilon(b_x + r_x \odot \cos(\vartheta - \tau)) \prod_{\jmath\in\mathcal{J}} J_K(\tau_\jmath)\, d\tau.$$

For each coordinate, $J_K(-t) = J_K(t)$: both sine factors in its defining ratio change sign, and the ratio is raised to the fourth power. The coordinatewise substitution $t = -\tau$ preserves $[-\pi,\pi]^{\mathcal{J}}$ and Lebesgue measure, and therefore changes $\cos(\vartheta-\tau)$ to $\cos(\vartheta+t)$ while leaving the product kernel unchanged. The last display is consequently the order-K tensor Jackson convolution in Definition 12 of $\bar{f}_\epsilon$ precomposed with the affine parametrization of Q_x, evaluated at ϑ.

5. Substituting the definition of S_x into the coefficient estimate gives

$$\sum_{\beta\in\operatorname{supp}(\mathcal{R}_{x,K})} |(\mathcal{R}_{x,K})_\beta| \le 2^{40K+20}(1+\epsilon^{-1})\sum_{\jmath\in\mathcal{J}} r_{\jmath,x}.$$

The constructed $P_{x,K}$ and $\mathcal{R}_{x,K}$ have the asserted physical degree, normalized degree, centered-evaluation identity, Jackson-convolution identity, and coefficient envelope.

□

The extraction step supplies the polynomial used by the estimator, and its centered form is the one matched to the factorial lift below. The dependence on $\sum_\jmath r_{\jmath,x}$ is the local scale that later becomes the Poisson noise scale after the pilot rectangle is chosen from the data.

For reference, we also isolate the tensor convolution operator used in these statements. This notation matches the product-kernel approximation scheme standard in Jackson-type polynomial approximation [DeVore and Lorentz, 1993].

Definition 19. [`synth_20`] (Tensor Jackson convolution operator $\mathcal{J}_K$) *For a positive integer K, let $J_K(t) = c_K\{\sin(Kt/2)/\sin(t/2)\}^4$ on $[-\pi,\pi]$, with the removable values filled in continuously and with c_K chosen so that $\int_{-\pi}^{\pi} J_K(t)\,dt = 1$. For a finite coordinate dimension D and a function $f:\mathbb{R}^D\to\mathbb{R}$, the order-$K$ tensor Jackson convolution is the operator $\mathcal{J}_K$ defined, for $\theta\in\mathbb{R}^D$, by*

$$(\mathcal{J}_K f)(\cos\theta) := \int_{[-\pi,\pi]^D} f\{\cos(\theta+t)\}\prod_{i=1}^{D} J_K(t_i)\,dt,$$

where cosine and addition are coordinatewise. Equivalently, the product kernel $\prod_{i=1}^D J_K(t_i)$ is normalized to have integral one over $[-\pi,\pi]^D$.

The approximation error is controlled through a weighted modulus along angular perturbations. The square-root boundary term in the main certificate comes from converting angular increments back to physical coordinates on a rectangle.

Lemma 3. [`lem:jackson-convolution-modulus`] (Tensor Jackson modulus) *Let D be a finite coordinate dimension, let K be a positive integer, and let $f:\mathbb{R}^D\to\mathbb{R}$. Let $L\ge 0$, let $a_i\ge 0$ and $b_i\ge 0$ for each $i\in[D]$, and fix an angle vector $\varphi\in\mathbb{R}^D$. Suppose:*

- ***(Kernel.)*** *The order-K tensor Jackson convolution is formed from the order-four Jackson kernel in Definition 12.*
- ***(Continuity.)*** *The function f is continuous on the normalized cube $[-1,1]^D$.*
- ***(Weighted increment modulus.)*** *For every shift $\tau\in\mathbb{R}^D$,*

$$|f\{\cos(\varphi-\tau)\} - f(\cos\varphi)| \le L\sum_{i=1}^{D}\left(a_i|\tau_i| + b_i\tau_i^2\right),$$

where cosine is applied coordinatewise.

Then
$$|(\mathcal{J}_K f)(\cos\varphi) - f(\cos\varphi)| \leq L \sum_{i=1}^{D} \left(\frac{32a_i}{K} + \frac{64b_i}{K^2} \right).$$

Proof of Lemma 3. 1. Let

$$\mathfrak{B}_D = [-\pi,\pi]^D \qquad \text{and} \qquad \mathfrak{k}_K^{(D)}(\tau) = \prod_{i=1}^{D} J_K(\tau_i), \quad \tau \in \mathfrak{B}_D.$$

For the positive integer K in the statement, J_K is the normalized order-four Jackson factor on $[-\pi,\pi]$, with removable values filled continuously. The raw fourth-power factor is nonnegative and has positive integral, so its normalized version satisfies $J_K(\rho) \geq 0$ for every $\rho \in [-\pi,\pi]$. The tensor product kernel is normalized to have unit mass:

$$\mathfrak{k}_K^{(D)}(\tau) \geq 0, \qquad \int_{\mathfrak{B}_D} \mathfrak{k}_K^{(D)}(\tau)\,d\tau = \prod_{i=1}^{D} \int_{-\pi}^{\pi} J_K(\rho)\,d\rho = 1.$$

The convolution is written in the shifted-variable form

$$(\mathcal{J}_K f)(\cos\varphi) = \int_{\mathfrak{B}_D} f\{\cos(\varphi-\tau)\}\mathfrak{k}_K^{(D)}(\tau)\,d\tau.$$

When $D = 0$, the product is the empty product and the sums below are empty sums, so these identities keep their displayed values. Continuity of f on the normalized cube, coordinatewise continuity of cosine, and compactness of $\mathfrak{B}_D$ give the integrability used below.

2. The preceding unit-mass identity gives

$$\begin{aligned}(\mathcal{J}_K f)(\cos\varphi) - f(\cos\varphi) &= \int_{\mathfrak{B}_D} f\{\cos(\varphi-\tau)\}\mathfrak{k}_K^{(D)}(\tau)\,d\tau - f(\cos\varphi)\int_{\mathfrak{B}_D} \mathfrak{k}_K^{(D)}(\tau)\,d\tau \\ &= \int_{\mathfrak{B}_D} \big[f\{\cos(\varphi-\tau)\} - f(\cos\varphi)\big]\mathfrak{k}_K^{(D)}(\tau)\,d\tau.\end{aligned}$$

The subtraction is justified by the compact-domain integrability noted above.

3. Taking absolute values, using nonnegativity of $\mathfrak{k}_K^{(D)}$, and applying the assumed weighted increment modulus pointwise yields

$$\begin{aligned}|(\mathcal{J}_K f)(\cos\varphi) - f(\cos\varphi)| &\leq \int_{\mathfrak{B}_D} |f\{\cos(\varphi-\tau)\} - f(\cos\varphi)|\,\mathfrak{k}_K^{(D)}(\tau)\,d\tau \\ &\leq \int_{\mathfrak{B}_D} L\sum_{i=1}^{D} \left(a_i|\tau_i| + b_i\tau_i^2\right)\mathfrak{k}_K^{(D)}(\tau)\,d\tau.\end{aligned}$$

4. Distributing the finite sum through the integral gives

$$\begin{aligned}&\int_{\mathfrak{B}_D} L\sum_{i=1}^{D} \left(a_i|\tau_i| + b_i\tau_i^2\right)\mathfrak{k}_K^{(D)}(\tau)\,d\tau \\ &\quad = L\sum_{i=1}^{D} \left[a_i\int_{\mathfrak{B}_D} |\tau_i|\mathfrak{k}_K^{(D)}(\tau)\,d\tau + b_i\int_{\mathfrak{B}_D} \tau_i^2\mathfrak{k}_K^{(D)}(\tau)\,d\tau\right].\end{aligned}$$

For each selected coordinate i, product integration and the one-dimensional unit mass give

$$\int_{\mathfrak{B}_D} |\tau_i| \mathfrak{k}_K^{(D)}(\tau)\, d\tau = \int_{-\pi}^{\pi} |\rho| J_K(\rho)\, d\rho$$

and

$$\int_{\mathfrak{B}_D} \tau_i^2 \mathfrak{k}_K^{(D)}(\tau)\, d\tau = \int_{-\pi}^{\pi} \rho^2 J_K(\rho)\, d\rho.$$

5. Write the raw kernel and its mass as

$$\widetilde{J}_K(\rho) = \begin{cases} K^4, & \sin(\rho/2) = 0, \\ \left\{ \dfrac{\sin(K\rho/2)}{\sin(\rho/2)} \right\}^4, & \sin(\rho/2) \neq 0, \end{cases} \qquad \mathfrak{m}_K = \int_{-\pi}^{\pi} \widetilde{J}_K(\rho)\, d\rho,$$

so that $J_K(\rho) = \widetilde{J}_K(\rho)/\mathfrak{m}_K$. Let U_m denote the Chebyshev polynomial of the second kind. The removable-value convention and the identity

$$U_{K-1}\{\cos(\rho/2)\} = \frac{\sin(K\rho/2)}{\sin(\rho/2)}$$

away from the removable points give

$$\widetilde{J}_K(\rho) = U_{K-1}\{\cos(\rho/2)\}^4.$$

The square expansion

$$U_{K-1}\{\cos(\rho/2)\}^2 = K + 2 \sum_{q \in \{0,\ldots,K-2\}} (K-1-q) \cos\{(q+1)\rho\}$$

follows from the Chebyshev recurrence, with the displayed sum interpreted as an empty sum when $K = 1$. Using

$$\int_{-\pi}^{\pi} \cos(r\rho)\, d\rho = 0, \qquad \int_{-\pi}^{\pi} \cos(r\rho)\cos(s\rho)\, d\rho = \begin{cases} \pi, & r = s, \\ 0, & r \neq s, \end{cases}$$

for positive integers r, s, one obtains

$$\begin{aligned} \mathfrak{m}_K &= \int_{-\pi}^{\pi} \left[K + 2 \sum_{q \in \{0,\ldots,K-2\}} (K-1-q) \cos\{(q+1)\rho\} \right]^2 d\rho \\ &= 2\pi K^2 + 4\pi \sum_{q \in \{0,\ldots,K-2\}} (K-1-q)^2 \\ &= 2\pi K^2 + \frac{2\pi}{3}(K-1)K(2K-1) = \frac{2\pi}{3} K(2K^2+1). \end{aligned}$$

In particular,

$$\mathfrak{m}_K \geq \frac{4\pi}{3} K^3 > 0.$$

6. The same Chebyshev-square expansion and the vanishing of all positive-frequency cosine integrals give

$$\int_{-\pi}^{\pi} U_{K-1}\{\cos(\rho/2)\}^2\,d\rho = \int_{-\pi}^{\pi}\left[K + 2\sum_{q\in\{0,\dots,K-2\}}(K-1-q)\cos\{(q+1)\rho\}\right]d\rho = 2\pi K.$$

For $\rho \in [-\pi,\pi]$,

$$\rho^2 \le \pi^2\sin^2(\rho/2).$$

If $\sin(\rho/2)=0$, then $\rho=0$ on this interval and

$$\rho^2\widetilde{J}_K(\rho) \le \pi^2 U_{K-1}\{\cos(\rho/2)\}^2.$$

If $\sin(\rho/2)\neq 0$, put

$$U_{K-1}\{\cos(\rho/2)\} = \frac{s_K(\rho)}{d(\rho)}, \qquad s_K(\rho)=\sin(K\rho/2), \qquad d(\rho)=\sin(\rho/2).$$

Since $s_K(\rho)^2 \le 1$,

$$\begin{aligned}
\rho^2\widetilde{J}_K(\rho) &= \rho^2\left\{\frac{s_K(\rho)}{d(\rho)}\right\}^4 \\
&= \rho^2\left\{\frac{s_K(\rho)}{d(\rho)}\right\}^2\left\{\frac{s_K(\rho)}{d(\rho)}\right\}^2 \\
&\le \pi^2 s_K(\rho)^2\left\{\frac{s_K(\rho)}{d(\rho)}\right\}^2 \le \pi^2\left\{\frac{s_K(\rho)}{d(\rho)}\right\}^2 = \pi^2 U_{K-1}\{\cos(\rho/2)\}^2.
\end{aligned}$$

After integration and normalization,

$$\begin{aligned}
\int_{-\pi}^{\pi}\rho^2 J_K(\rho)\,d\rho &= \frac{1}{\mathfrak{m}_K}\int_{-\pi}^{\pi}\rho^2\widetilde{J}_K(\rho)\,d\rho \\
&\le \frac{\pi^2}{\mathfrak{m}_K}\int_{-\pi}^{\pi} U_{K-1}\{\cos(\rho/2)\}^2\,d\rho \\
&\le \frac{2\pi^3 K}{(4\pi/3)K^3} = \frac{3\pi^2}{2K^2} \le \frac{24}{K^2} \le \frac{64}{K^2}.
\end{aligned}$$

7. The elementary inequality

$$0 \le (K|\rho|-2)^2$$

is equivalent, since $K>0$, to

$$|\rho| \le \frac{K}{4}\rho^2 + \frac{1}{K}.$$

Multiplying by $J_K(\rho)\ge 0$, integrating, and using the preceding second-moment estimate together with $\int_{-\pi}^{\pi} J_K(\rho)\,d\rho = 1$, gives

$$\begin{aligned}
\int_{-\pi}^{\pi}|\rho| J_K(\rho)\,d\rho &\le \frac{K}{4}\int_{-\pi}^{\pi}\rho^2 J_K(\rho)\,d\rho + \frac{1}{K}\int_{-\pi}^{\pi} J_K(\rho)\,d\rho \\
&\le \frac{K}{4}\frac{24}{K^2} + \frac{1}{K} = \frac{7}{K} \le \frac{32}{K}.
\end{aligned}$$

8. Combining the integral bound from the previous steps with the tensor moment reductions gives

$$|(\mathcal{J}_K f)(\cos\varphi) - f(\cos\varphi)| \le L \sum_{i=1}^{D} \left[a_i \frac{32}{K} + b_i \frac{64}{K^2} \right] = L \sum_{i=1}^{D} \left(\frac{32a_i}{K} + \frac{64b_i}{K^2} \right).$$

Here $L \ge 0$, $a_i \ge 0$, and $b_i \ge 0$ justify multiplying the one-dimensional moment inequalities and summing them coordinatewise. This is the asserted bound.

□

Combining the coefficient envelope with the angular modulus gives the single polynomial certificate used in the upper-bound proof. It applies uniformly to nonnegative rectangles with positive radii, including the pilot-local rectangles that may meet coordinate faces or the vertex of the overlap cone.

Lemma 4. [`lem:simultaneous-jackson-certificate`] (Simultaneous Jackson certificate) *There is a universal constant $A > 0$ such that, for every overlap parameter ϵ with $0 < \epsilon < 1/2$, there is a constant $C_\epsilon > 0$ with the following property. Let:*

- ***(Degree.)*** *$K \ge 2$.*
- ***(Rectangle.)*** *$Q_x = \prod_{\jmath\in\mathcal{J}}[\ell_{\jmath,x}, u_{\jmath,x}]$ have $\ell_{\jmath,x} \le u_{\jmath,x}$, $\ell_{\jmath,x} \ge 0$, and positive radii*

$$r_{\jmath,x} = \frac{u_{\jmath,x} - \ell_{\jmath,x}}{2} > 0$$

for every $\jmath \in \mathcal{J}$.

Let $P_{x,K}$ be the tensor Jackson polynomial associated with ϵ, K, and Q_x in Definition 12. Then the same polynomial satisfies all of the following conclusions:

$$\deg_\jmath(P_{x,K}) \le 2(K-1) \qquad \text{for every } \jmath \in \mathcal{J}.$$

For every $v \in Q_x$,

$$\left|P_{x,K}(v) - \bar{f}_\epsilon(v)\right| \le C_\epsilon \sum_{\jmath\in\mathcal{J}} \left\{ \frac{\sqrt{(v_\jmath - \ell_{\jmath,x})(u_{\jmath,x} - v_\jmath)}}{K} + \frac{r_{\jmath,x}}{K^2} \right\}.$$

There are a finite multi-index set $\mathcal{A}_x$ and coefficients $b_{x,\alpha}$ such that, for every v,

$$P_{x,K}(v) - \bar{f}_\epsilon(b_x) = \sum_{\alpha\in\mathcal{A}_x} b_{x,\alpha} \prod_{\jmath\in\mathcal{J}} \left(\frac{v_\jmath - b_{\jmath,x}}{r_{\jmath,x}} \right)^{\alpha_\jmath},$$

with

$$\alpha_\jmath \le 2(K-1) \qquad \text{for every } \alpha \in \mathcal{A}_x \text{ and every } \jmath \in \mathcal{J},$$

and

$$\sum_{\alpha\in\mathcal{A}_x} |b_{x,\alpha}| \le A^K (1 + \epsilon^{-1}) \sum_{\jmath\in\mathcal{J}} r_{\jmath,x}.$$

Proof of Lemma 4. Take

$$A = 2^{60}.$$

Fix $\epsilon \in (0, 1/2)$, and set

$$C_\epsilon = 32(1 + \epsilon^{-1}).$$

Let $K \geq 2$ and let Q_x satisfy the displayed hypotheses. Applying Lemma 2 to this rectangle gives the tensor Jackson polynomial $P_{x,K}$ and a normalized centered polynomial $\mathcal{R}_{x,K}$. The same result gives, first,

$$\deg_\jmath(P_{x,K}) \leq 2(K-1) \qquad (\jmath \in \mathcal{J}).$$

Fix $v \in Q_x$. Choose a bijection $\eta : \mathcal{J} \to \{1, 2, 3, 4\}$, and write

$$c_\omega = b_{\eta^{-1}(\omega),x}, \qquad \rho_\omega = r_{\eta^{-1}(\omega),x}, \qquad y_\omega = v_{\eta^{-1}(\omega)} \qquad (\omega = 1, \ldots, 4).$$

Define the normalized point and angle vector by

$$z_\omega = \frac{y_\omega - c_\omega}{\rho_\omega}, \qquad \varphi_\omega = \arccos z_\omega.$$

Since $v \in Q_x$ and every radius is positive, $z \in [-1, 1]^4$ and

$$v_\jmath = b_{\jmath,x} + r_{\jmath,x} \cos \varphi_{\eta(\jmath)} \qquad (\jmath \in \mathcal{J}).$$

On $[-1, 1]^4$, set

$$F_x(w) = \bar{f}_\epsilon\Big(\big\{b_{\jmath,x} + r_{\jmath,x} w_{\eta(\jmath)}\big\}_{\jmath \in \mathcal{J}}\Big).$$

The lower-endpoint hypothesis gives $b_{\jmath,x} - r_{\jmath,x} = \ell_{\jmath,x} \geq 0$, so the affine image of $[-1, 1]^4$ lies in $\mathbb{R}^4_+$. The global Lipschitz property in Proposition 1 therefore implies that F_x is continuous on $[-1, 1]^4$ and that, for all $w, w' \in [-1, 1]^4$,

$$|F_x(w) - F_x(w')| \leq (1 + \epsilon^{-1}) \sum_{\omega=1}^{4} \rho_\omega |w_\omega - w'_\omega|.$$

For every shift $\tau \in \mathbb{R}^4$, the elementary inequality

$$|\cos(\varphi_\omega - \tau_\omega) - \cos \varphi_\omega| \leq |\sin \varphi_\omega|\, |\tau_\omega| + \frac{\tau_\omega^2}{2}$$

therefore gives

$$|F_x\{\cos(\varphi - \tau)\} - F_x(\cos \varphi)| \leq (1 + \epsilon^{-1}) \sum_{\omega=1}^{4} \left(\rho_\omega |\sin \varphi_\omega|\, |\tau_\omega| + \frac{\rho_\omega}{2} \tau_\omega^2\right).$$

Applying Lemma 3 with

$$L = 1 + \epsilon^{-1}, \qquad a_\omega = \rho_\omega |\sin \varphi_\omega|, \qquad b_\omega = \frac{\rho_\omega}{2}$$

yields

$$|P_{x,K}(v) - \bar{f}_\epsilon(v)| \leq (1 + \epsilon^{-1}) \sum_{\omega=1}^{4} \left(\frac{32\rho_\omega |\sin \varphi_\omega|}{K} + \frac{32\rho_\omega}{K^2}\right),$$

where the evaluation identity for $P_{x,K}$ is the Jackson-convolution identity supplied by Lemma 2. Finally, for each $\jmath \in \mathcal{J}$,

$$r_{\jmath,x}|\sin\varphi_{\eta(\jmath)}| = \sqrt{(v_\jmath - \ell_{\jmath,x})(u_{\jmath,x} - v_\jmath)},$$

because

$$(v_\jmath - \ell_{\jmath,x})(u_{\jmath,x} - v_\jmath) = r_{\jmath,x}^2 \left\{1 - \left(\frac{v_\jmath - b_{\jmath,x}}{r_{\jmath,x}}\right)^2\right\}.$$

Reindexing the preceding sum over $\mathcal{J}$ and using $C_\epsilon = 32(1+\epsilon^{-1})$ gives the asserted pointwise bound.

It remains to record the centered expansion and its coefficient envelope. By Lemma 2, for every four-coordinate vector v,

$$P_{x,K}(v) - \bar{f}_\epsilon(b_x) = \mathcal{R}_{x,K}\left(\left\{\frac{v_\jmath - b_{\jmath,x}}{r_{\jmath,x}}\right\}_{\jmath\in\mathcal{J}}\right).$$

Let $\mathcal{A}_x$ be the image of $\operatorname{supp}(\mathcal{R}_{x,K})$ under the coordinate reindexing from $\{1,2,3,4\}$ to $\mathcal{J}$, and let $b_{x,\alpha}$ be the corresponding reindexed coefficient. Then the preceding display becomes

$$P_{x,K}(v) - \bar{f}_\epsilon(b_x) = \sum_{\alpha\in\mathcal{A}_x} b_{x,\alpha} \prod_{\jmath\in\mathcal{J}} \left(\frac{v_\jmath - b_{\jmath,x}}{r_{\jmath,x}}\right)^{\alpha_\jmath}.$$

The normalized coordinate-degree conclusion of Lemma 2 gives

$$\alpha_\jmath \le 2(K-1) \qquad (\alpha \in \mathcal{A}_x,\ \jmath \in \mathcal{J}),$$

and reindexing preserves the coefficient one-norm, so its coefficient envelope gives

$$\sum_{\alpha\in\mathcal{A}_x} |b_{x,\alpha}| \le 2^{40K+20}(1+\epsilon^{-1}) \sum_{\jmath\in\mathcal{J}} r_{\jmath,x}.$$

Since $K \ge 2$,

$$40K + 20 \le 60K, \qquad 2^{40K+20} \le (2^{60})^K = A^K.$$

Combining the last two displays proves the stated coefficient bound. ☐

Lemma 4 is the analytic bridge between approximation and estimation. The pointwise bound controls the local bias of $P_{x,K}$ on Q_x, while the centered coefficient expansion controls the moments of the factorial estimator after replacing monomials by centered Poisson factorial polynomials.

We next record the pilot calculations that make the random rectangles usable. The good-pilot event is self-normalized: its radius is computed from the observed pilot count itself, so the event has the same scale in small and large cells.

Lemma 5. [`lem:canonical-pilot-bad-moment`] (Bad pilot moment bound) *Fix the four-coordinate set $\mathcal{J} = \{0,1\} \times \{0,1\}$. Let $q = (q_\jmath)_{\jmath\in\mathcal{J}}$ have nonnegative coordinates, let $m > 0$, let $L \ge 1$, and let $t \in \{1,2,4\}$. Suppose the pilot counts $(N'_\jmath)_{\jmath\in\mathcal{J}}$ are independent with*

$$N'_\jmath \sim \operatorname{Pois}(mq_\jmath), \qquad \jmath \in \mathcal{J}.$$

With $H_0 = 1024$, define the normalized aggregate pilot score

$$\Gamma(N') = \sum_{\jmath \in \mathcal{J}} \left[\left| \frac{N'_\jmath}{m} - q_\jmath \right| + H_0 \left\{ \sqrt{\frac{N'_\jmath}{m} \frac{L}{m}} + \frac{L}{m} \right\} \right],$$

and define the self-normalized bad-pilot event

$$\mathsf{G}^c = \left\{ \exists \jmath \in \mathcal{J} : \left| \frac{N'_\jmath}{m} - q_\jmath \right| > \frac{H_0}{4} \left\{ \sqrt{\frac{N'_\jmath}{m} \frac{L}{m}} + \frac{L}{m} \right\} \right\}.$$

Then, with $C_t = 10^{80(t+1)}$,

$$E\big[\Gamma(N')^t \mathbf{1}_{\mathsf{G}^c}\big] \le C_t e^{-20L} \left\{ \sqrt{\frac{\left(\sum_{\jmath \in \mathcal{J}} q_\jmath\right) L}{m}} + \frac{L}{m} \right\}^t.$$

Proof of Lemma 5. 1. For each $\jmath \in \mathcal{J}$, define

$$\eta_\jmath = m q_\jmath, \qquad \delta_\jmath = \left| N'_\jmath - \eta_\jmath \right|, \qquad \rho_\jmath = \sqrt{N'_\jmath L} + L, \qquad \sigma_\jmath = \delta_\jmath + 1024 \rho_\jmath,$$

and put

$$\Sigma = \sum_{\jmath \in \mathcal{J}} \sigma_\jmath.$$

Since $m > 0$,

$$\left| \frac{N'_\jmath}{m} - q_\jmath \right| = \frac{\delta_\jmath}{m}, \qquad \sqrt{\frac{N'_\jmath}{m} \frac{L}{m}} + \frac{L}{m} = \frac{\rho_\jmath}{m}.$$

Thus

$$\Gamma(N') = \frac{\Sigma}{m}, \qquad \mathsf{G}^c = \bigcup_{\jmath \in \mathcal{J}} \{\delta_\jmath > 256 \rho_\jmath\},$$

and hence

$$E\big[\Gamma(N')^t \mathbf{1}_{\mathsf{G}^c}\big] = m^{-t} E\Big[\Sigma^t \mathbf{1}_{\cup_{\jmath \in \mathcal{J}} \{\delta_\jmath > 256 \rho_\jmath\}}\Big].$$

2. We first record the scalar estimate used in the product step. Let $\mathsf{W} \sim \mathrm{Pois}(\Lambda)$, where $\Lambda \ge 0$, and define

$$\delta_\Lambda(\mathsf{W}) = |\mathsf{W} - \Lambda|, \qquad \rho_L(\mathsf{W}) = \sqrt{\mathsf{W} L} + L, \qquad \sigma_{\Lambda,L}(\mathsf{W}) = \delta_\Lambda(\mathsf{W}) + 1024 \rho_L(\mathsf{W}),$$

$$\upsilon_{\Lambda,L} = \sqrt{\Lambda L} + L, \qquad \mathcal{B}_{\Lambda,L} = \{\delta_\Lambda(\mathsf{W}) > 256 \rho_L(\mathsf{W})\}.$$

For every real ξ,

$$E \exp\{\xi(\mathsf{W} - \Lambda)\} = \exp\{\Lambda(e^\xi - 1 - \xi)\}.$$

The Chernoff–Bennett bounds obtained from this identity give, for every $\zeta_{\text{tail}} \ge 0$,

$$\Pr\{\mathsf{W} - \Lambda > \sqrt{2\Lambda \zeta_{\text{tail}}} + 2\zeta_{\text{tail}}\} \le e^{-\zeta_{\text{tail}}}, \qquad \Pr\{\Lambda - \mathsf{W} > \sqrt{2\Lambda \zeta_{\text{tail}}}\} \le e^{-\zeta_{\text{tail}}}.$$

With $\zeta_{\text{tail}} = 80L$,

$$\mathcal{B}_{\Lambda,L} \subseteq \{\mathsf{W} - \Lambda > \sqrt{160 \Lambda L} + 160L\} \cup \{\Lambda - \mathsf{W} > \sqrt{160 \Lambda L}\}.$$

Indeed, on $\{\Lambda \le \mathsf{W}\}$,

$$\sqrt{160\Lambda L} \le 16\sqrt{\mathsf{W}L},$$

so the bad-event inequality implies the displayed upper-tail event. On $\{\mathsf{W} < \Lambda\}$, the inequality

$$\Lambda - \mathsf{W} > 256(\sqrt{\mathsf{W}L} + L)$$

implies

$$(\Lambda - \mathsf{W})^2 - 160L(\Lambda - \mathsf{W}) - 160\mathsf{W}L > 0,$$

and therefore $(\Lambda - \mathsf{W})^2 > 160\Lambda L$, which is the displayed lower-tail event. Consequently

$$\Pr(\mathcal{B}_{\Lambda,L}) \le 2e^{-80L}.$$

Next,

$$\sqrt{\mathsf{W}L} \le \sqrt{\Lambda L} + \sqrt{\delta_\Lambda(\mathsf{W})L} \le \sqrt{\Lambda L} + \frac{\delta_\Lambda(\mathsf{W}) + L}{2},$$

and hence

$$\sigma_{\Lambda,L}(\mathsf{W}) \le 1536\{\upsilon_{\Lambda,L} + \delta_\Lambda(\mathsf{W})\}.$$

Since $L \ge 1$, $\upsilon_{\Lambda,L} \ge 1$. Set

$$\xi_{\Lambda,L} = \frac{1}{2\upsilon_{\Lambda,L}}.$$

Then $|\xi_{\Lambda,L}| \le 1$ and $\Lambda\xi_{\Lambda,L}^2 \le 1/4$. The preceding deterministic bound gives

$$\exp\left\{\frac{\sigma_{\Lambda,L}(\mathsf{W})}{4096\upsilon_{\Lambda,L}}\right\} \le e^{1/2}\left[e^{\xi_{\Lambda,L}(\mathsf{W}-\Lambda)} + e^{-\xi_{\Lambda,L}(\mathsf{W}-\Lambda)}\right].$$

Using the moment-generating identity and $|e^r - 1 - r| \le r^2$ for $|r| \le 1$,

$$E\exp\left\{\frac{\sigma_{\Lambda,L}(\mathsf{W})}{4096\upsilon_{\Lambda,L}}\right\} \le e^{1/2}(e^{1/4} + e^{1/4}) \le 8.$$

The elementary exponential domination of powers then yields, for every integer s with $0 \le s \le 8$,

$$E\sigma_{\Lambda,L}(\mathsf{W})^s \le 10^{4s+4}\upsilon_{\Lambda,L}^s.$$

For $t \in \{1, 2, 4\}$, define

$$\zeta_{\Lambda,L,t} = e^{-40L}\upsilon_{\Lambda,L}^{-t}.$$

With

$$X_{\Lambda,L,t}(\mathsf{W}) = \sigma_{\Lambda,L}(\mathsf{W})^t,$$

the pointwise Young inequality gives

$$\mathbf{1}_{\mathcal{B}_{\Lambda,L}} X_{\Lambda,L,t} \le \frac{\zeta_{\Lambda,L,t}X_{\Lambda,L,t}^2 + \zeta_{\Lambda,L,t}^{-1}\mathbf{1}_{\mathcal{B}_{\Lambda,L}}}{2}.$$

Combining the moment bound with $s = 2t$ and the bad-event probability bound,

$$E\left[\sigma_{\Lambda,L}(\mathsf{W})^t\mathbf{1}_{\mathcal{B}_{\Lambda,L}}\right] \le \frac{10^{8t+4} + 2}{2}\,e^{-40L}\upsilon_{\Lambda,L}^t \le 10^{8t+8}e^{-40L}\upsilon_{\Lambda,L}^t.$$

3. Now let $\mathcal{I}$ be a finite coordinate set with $|\mathcal{I}| \le 4$, and let $(\mathsf{W}_\iota)_{\iota\in\mathcal{I}}$ be independent with $\mathsf{W}_\iota \sim \mathrm{Pois}(\Lambda_\iota)$. Define

$$\delta_\iota = |\mathsf{W}_\iota - \Lambda_\iota|, \qquad \rho_\iota = \sqrt{\mathsf{W}_\iota L} + L, \qquad \sigma_\iota = \delta_\iota + 1024\rho_\iota,$$

$$\mathcal{B}_\iota = \{\delta_\iota > 256\rho_\iota\}, \qquad \Sigma_{\mathcal{I}} = \sum_{\iota\in\mathcal{I}} \sigma_\iota, \qquad \Xi_{\mathcal{I},L} = \sqrt{\left(\sum_{\iota\in\mathcal{I}} \Lambda_\iota\right) L} + L.$$

If $\mathcal{I} = \varnothing$, the corresponding integral is zero. Otherwise, since each $\sigma_\iota \ge 0$,

$$\Sigma_{\mathcal{I}}^t \le |\mathcal{I}|^{t-1} \sum_{\iota\in\mathcal{I}} \sigma_\iota^t,$$

for $t \in \{1,2,4\}$, and therefore

$$E\big[\Sigma_{\mathcal{I}}^t \mathbf{1}_{\cup_{\iota'\in\mathcal{I}}\mathcal{B}_{\iota'}}\big] \le |\mathcal{I}|^{t-1} \sum_{\iota',\iota\in\mathcal{I}} E\big[\sigma_\iota^t \mathbf{1}_{\mathcal{B}_{\iota'}}\big].$$

When $\iota = \iota'$, the scalar bound from step 2 gives

$$E\big[\sigma_\iota^t \mathbf{1}_{\mathcal{B}_\iota}\big] \le 10^{8t+8} e^{-40L} \left(\sqrt{\Lambda_\iota L} + L\right)^t.$$

When $\iota \ne \iota'$, independence factors the expectation and the scalar moment and probability bounds give

$$E\big[\sigma_\iota^t \mathbf{1}_{\mathcal{B}_{\iota'}}\big] = E\sigma_\iota^t \,\Pr(\mathcal{B}_{\iota'}) \le 2\cdot 10^{4t+4} e^{-40L} \left(\sqrt{\Lambda_\iota L} + L\right)^t.$$

Thus, with

$$\mathfrak{c}_t = 10^{8t+8} + 2\cdot 10^{4t+4},$$

$$E\big[\Sigma_{\mathcal{I}}^t \mathbf{1}_{\cup_{\iota'\in\mathcal{I}}\mathcal{B}_{\iota'}}\big] \le |\mathcal{I}|^t \mathfrak{c}_t e^{-40L} \sum_{\iota\in\mathcal{I}} \left(\sqrt{\Lambda_\iota L} + L\right)^t.$$

Cauchy–Schwarz gives

$$\sum_{\iota\in\mathcal{I}} \sqrt{\Lambda_\iota L} \le |\mathcal{I}| \sqrt{\left(\sum_{\iota\in\mathcal{I}} \Lambda_\iota\right) L},$$

and hence

$$\sum_{\iota\in\mathcal{I}} \left(\sqrt{\Lambda_\iota L} + L\right)^t \le |\mathcal{I}|^{t+1} \Xi_{\mathcal{I},L}^t.$$

Since $|\mathcal{I}| \le 4$,

$$|\mathcal{I}|^{2t+1} \mathfrak{c}_t \le 10^{16(4+1)(t+1)} = 10^{80(t+1)}.$$

Also $e^{-40L} \le e^{-20L}$. Therefore

$$E\big[\Sigma_{\mathcal{I}}^t \mathbf{1}_{\cup_{\iota'\in\mathcal{I}}\mathcal{B}_{\iota'}}\big] \le 10^{80(t+1)} e^{-20L} \left\{ \sqrt{\left(\sum_{\iota\in\mathcal{I}} \Lambda_\iota\right) L} + L \right\}^t.$$

4. Apply the preceding product bound with

$$\mathcal{I} = \mathcal{J}, \qquad \Lambda_\jmath = mq_\jmath.$$

The product-Poisson pilot law supplies the displayed coordinate laws and independence, and $|\mathcal{J}| = 4$. Hence

$$E\left[\Sigma^t \mathbf{1}_{\cup_{\jmath\in\mathcal{J}}\{\delta_\jmath > 256\rho_\jmath\}}\right] \le 10^{80(t+1)} e^{-20L} \left\{ \sqrt{\left(\sum_{\jmath\in\mathcal{J}} m q_\jmath\right) L} + L \right\}^t .$$

Combining this estimate with step 1 and using $m > 0$,

$$m^{-1}\left\{ \sqrt{m\left(\sum_{\jmath\in\mathcal{J}} q_\jmath\right) L} + L \right\} = \sqrt{\frac{\left(\sum_{\jmath\in\mathcal{J}} q_\jmath\right) L}{m}} + \frac{L}{m}.$$

Therefore

$$E\left[\Gamma(N')^t \mathbf{1}_{\mathsf{G}^c}\right] \le 10^{80(t+1)} e^{-20L} \left\{ \sqrt{\frac{\left(\sum_{\jmath\in\mathcal{J}} q_\jmath\right) L}{m}} + \frac{L}{m} \right\}^t .$$

Since $C_t = 10^{80(t+1)}$, this is the asserted bound. □

The exponential factor in Lemma 5 is paired with the logarithmic choice $L_d = \log(ed)$ from Algorithm 1. This pairing makes the contribution from bad pilot rectangles summable across cells at the rate used in Theorem 1.

Definition 20. [`synth_12`] (Good-pilot event) *Fix a cell x, four-cell mass vector $q_x = (q_{\jmath,x})_{\jmath\in\mathcal{J}}$, Poisson scale $m > 0$, alphabet logarithm $L_d = \log(ed)$, pilot-radius constant H_0, and pilot counts $(N'_{\jmath,x})_{\jmath\in\mathcal{J}}$. Put $c_{\jmath,x} = N'_{\jmath,x}/m$ and $h_{\jmath,x} = H_0\{\sqrt{c_{\jmath,x}L_d/m} + L_d/m\}$. The coordinatewise good-pilot event used for the pilot rectangle is*

$$\mathsf{G}_x := \bigcap_{\jmath\in\mathcal{J}} \left\{ \left| \frac{N'_{\jmath,x}}{m} - q_{\jmath,x} \right| \le \frac{h_{\jmath,x}}{4} \right\}.$$

On G_x, the pilot rectangle radius is comparable to the local Poisson scale. The next lemma expresses that comparison in the aggregate form needed by the coefficient envelope.

Lemma 6. [`lem:good-pilot-radius-sum`] (Good pilot radius bound) *Fix $m > 0$, an integer $d \ge 1$, pilot counts $N'_{\jmath,x}$ for $\jmath$ in the four-coordinate set $\mathcal{J}$, and coordinate intensities $q_\jmath \ge 0$. Let $L_d = \log(ed)$, set $p_x = \sum_{\jmath\in\mathcal{J}} q_\jmath$, and let $Q_x = \prod_{\jmath\in\mathcal{J}}[\ell_{\jmath,x}, u_{\jmath,x}]$ be the canonical pilot rectangle of Algorithm 1, with radii*

$$r_{\jmath,x} = \frac{u_{\jmath,x} - \ell_{\jmath,x}}{2}.$$

Assume that the good-pilot event

$$\mathsf{G}_x = \bigcap_{\jmath\in\mathcal{J}} \left\{ |c_{\jmath,x} - q_\jmath| \le \frac{h_{\jmath,x}}{4} \right\}, \qquad c_{\jmath,x} = \frac{N'_{\jmath,x}}{m},$$

holds, where $h_{\jmath,x}$ is the canonical pilot half-width in Algorithm 1. Then

$$\sum_{\jmath\in\mathcal{J}} r_{\jmath,x} \le 2\cdot 10^6 \left\{ \sqrt{\frac{p_x L_d}{m}} + \frac{L_d}{m} \right\}.$$

Proof of Lemma 6. Fix x and set

$$\zeta = \frac{L_d}{m}.$$

Since $d \geq 1$, $ed \geq e > 1$, so $L_d = \log(ed) > 0$. Hence $\zeta > 0$. For each $\jmath \in \mathcal{J}$, write

$$\sigma_{\jmath,x} = \sqrt{c_{\jmath,x}\zeta}.$$

For the canonical tuning used in this lemma, the pilot half-width is calibrated with $H_0 = 1024$. Hence

$$h_{\jmath,x} = 1024(\sigma_{\jmath,x} + \zeta).$$

Also $\sigma_{\jmath,x} \geq 0$, and

$$\sigma_{\jmath,x} \leq \frac{c_{\jmath,x}}{512} + 128\zeta$$

because both sides are nonnegative and

$$\left(\frac{c_{\jmath,x}}{512} + 128\zeta\right)^2 - c_{\jmath,x}\zeta = \left(\frac{c_{\jmath,x}}{512} - 128\zeta\right)^2 \geq 0.$$

The good-pilot event gives

$$c_{\jmath,x} \leq q_\jmath + \frac{h_{\jmath,x}}{4}.$$

Substituting the displayed formula for $h_{\jmath,x}$ and then using the preceding bound on $\sigma_{\jmath,x}$,

$$c_{\jmath,x} \leq q_\jmath + 256\sigma_{\jmath,x} + 256\zeta \leq q_\jmath + \frac{c_{\jmath,x}}{2} + 33024\zeta.$$

Thus

$$c_{\jmath,x} \leq 2q_\jmath + 66048\zeta. \tag{1}$$

Define

$$\rho_{\jmath,x} = \sqrt{q_\jmath\zeta}.$$

Since $q_\jmath \geq 0$ and $\zeta > 0$, we have $\rho_{\jmath,x}^2 = q_\jmath\zeta$. From Equation (1),

$$\sigma_{\jmath,x}^2 = c_{\jmath,x}\zeta \leq (2q_\jmath + 66048\zeta)\zeta.$$

Moreover,

$$(2\rho_{\jmath,x} + 300\zeta)^2 - (2q_\jmath + 66048\zeta)\zeta = 2q_\jmath\zeta + 1200\rho_{\jmath,x}\zeta + 23952\zeta^2 \geq 0.$$

Taking square roots gives

$$\sigma_{\jmath,x} \leq 2\sqrt{q_\jmath\zeta} + 300\zeta. \tag{2}$$

For the zero-truncated pilot interval,

$$r_{\jmath,x} = \frac{c_{\jmath,x} + h_{\jmath,x} - \max\{0, c_{\jmath,x} - h_{\jmath,x}\}}{2}.$$

If $c_{\jmath,x} \leq h_{\jmath,x}$, then $r_{\jmath,x} = (c_{\jmath,x} + h_{\jmath,x})/2 \leq h_{\jmath,x}$. If $c_{\jmath,x} > h_{\jmath,x}$, then $r_{\jmath,x} = h_{\jmath,x}$. Therefore

$$r_{\jmath,x} \leq h_{\jmath,x}.$$

Combining this with the formula for $h_{\jmath,x}$ and Equation (2),

$$r_{\jmath,x} \le 1024(\sigma_{\jmath,x} + \zeta) \le 2048\sqrt{q_\jmath \zeta} + 308224\zeta. \tag{3}$$

Summing Equation (3) over the four coordinates,

$$\sum_{\jmath \in \mathcal{J}} r_{\jmath,x} \le 2048 \sum_{\jmath \in \mathcal{J}} \sqrt{q_\jmath \zeta} + 4 \cdot 308224\, \zeta.$$

By Cauchy's inequality on the four-point set $\mathcal{J}$,

$$\left(\sum_{\jmath \in \mathcal{J}} \sqrt{q_\jmath \zeta} \right)^2 \le 4 \sum_{\jmath \in \mathcal{J}} \left(\sqrt{q_\jmath \zeta}\right)^2 = 4 \left(\sum_{\jmath \in \mathcal{J}} q_\jmath \right) \zeta.$$

All terms are nonnegative, so

$$\sum_{\jmath \in \mathcal{J}} \sqrt{q_\jmath \zeta} \le 2 \sqrt{\left(\sum_{\jmath \in \mathcal{J}} q_\jmath \right) \zeta}.$$

Consequently,

$$\sum_{\jmath \in \mathcal{J}} r_{\jmath,x} \le 4096 \sqrt{\left(\sum_{\jmath \in \mathcal{J}} q_\jmath \right) \zeta} + 1232896\, \zeta \le 2 \cdot 10^6 \left\{ \sqrt{\left(\sum_{\jmath \in \mathcal{J}} q_\jmath \right) \zeta} + \zeta \right\}.$$

Since $p_x = \sum_{\jmath \in \mathcal{J}} q_\jmath$ and $\zeta = L_d/m$, the desired bound follows:

$$\sum_{\jmath \in \mathcal{J}} r_{\jmath,x} \le 2 \cdot 10^6 \left\{ \sqrt{\frac{p_x L_d}{m}} + \frac{L_d}{m} \right\}.$$

□

The radius bound turns the abstract coefficient control into a stochastic bound indexed by p_x, the cell mass. The following normalized envelope is the version applied directly to the centered polynomial expansion inside the estimator.

Lemma 7. [`lem:jackson-normalized-coefficient-envelope`] (Normalized coefficient envelope) *Let $\epsilon > 0$ be the overlap parameter from Assumption 4, let $K \ge 2$ be an integer, and let $Q_x = \prod_{\jmath \in \mathcal{J}} [\ell_{\jmath,x}, u_{\jmath,x}]$ be a rectangle over the four outcome-arm coordinate set $\mathcal{J}$. Suppose that:*

- ***(Ordered endpoints.)*** *$\ell_{\jmath,x} \le u_{\jmath,x}$ for every $\jmath \in \mathcal{J}$.*
- ***(Nonnegative lower endpoints.)*** *$\ell_{\jmath,x} \ge 0$ for every $\jmath \in \mathcal{J}$.*
- ***(Positive radii.)*** *With*

$$b_{\jmath,x} = \frac{\ell_{\jmath,x} + u_{\jmath,x}}{2}, \qquad r_{\jmath,x} = \frac{u_{\jmath,x} - \ell_{\jmath,x}}{2},$$

  one has $r_{\jmath,x} > 0$ for every $\jmath \in \mathcal{J}$.

Let $b_x = (b_{\jmath,x})_{\jmath\in\mathcal{J}}$, let $r_x = (r_{\jmath,x})_{\jmath\in\mathcal{J}}$, and let $P_{x,K}$ be the Jackson tensor polynomial attached to Q_x by Definition 12. Define

$$\mathcal{R}_{x,K}(z) = P_{x,K}\big((b_{\jmath,x} + r_{\jmath,x}z_\jmath)_{\jmath\in\mathcal{J}}\big) - \bar{f}_\epsilon(b_x) = \sum_{\alpha\in\mathcal{A}_x} b_{x,\alpha} \prod_{\jmath\in\mathcal{J}} z_\jmath^{\alpha_\jmath},$$

where $\bar{f}_\epsilon$ is the global cell extension from Definition 8 and $\mathcal{A}_x = \operatorname{supp}(\mathcal{R}_{x,K})$. Then

$$\sum_{\alpha\in\mathcal{A}_x} |b_{x,\alpha}| \le (2^{60})^K(1+\epsilon^{-1}) \sum_{\jmath\in\mathcal{J}} r_{\jmath,x}.$$

Moreover, every $\alpha \in \mathcal{A}_x$ satisfies

$$\alpha_\jmath \le 2(K-1)$$

for every $\jmath \in \mathcal{J}$.

Proof of Lemma 7. 1. The ordered endpoints make Q_x a valid rectangle, and the remaining hypotheses give the lower-endpoint, positive-overlap, degree, and positive-radius conditions needed for Lemma 2. Apply that result to Q_x. Under the same positive-overlap and positive-radius conditions, the construction in Definition 12 identifies the extracted physical polynomial with $P_{x,K}$.

Fix a bijection

$$\iota_{\mathcal{J}} : \mathcal{J} \longrightarrow \{1,2,3,4\}.$$

Write the normalized polynomial supplied by Lemma 2 as

$$\mathcal{Q}_{x,K}(\zeta) = \sum_{\beta\in\mathcal{B}_x} \eta_\beta \prod_{\varpi=1}^{4} \zeta_\varpi^{\beta_\varpi}, \qquad \mathcal{B}_x = \operatorname{supp}(\mathcal{Q}_{x,K}) \subset \mathbb{N}^{\{1,2,3,4\}}, \qquad \eta_\beta = (\mathcal{Q}_{x,K})_\beta \quad (\beta\in\mathcal{B}_x).$$

Define the coordinate transport

$$\Phi : \mathbb{N}^{\{1,2,3,4\}} \longrightarrow \mathbb{N}^{\mathcal{J}}, \qquad (\Phi\beta)_\jmath = \beta_{\iota_{\mathcal{J}}(\jmath)},$$

and set

$$\mathcal{S}_x = \Phi(\mathcal{B}_x), \qquad \xi_{x,\alpha} = \eta_{\Phi^{-1}(\alpha)} \quad \text{for } \alpha\in\mathcal{S}_x.$$

Then, for every $v \in \mathbb{R}^{\mathcal{J}}$,

$$P_{x,K}(v) - \bar{f}_\epsilon(b_x) = \sum_{\alpha\in\mathcal{S}_x} \xi_{x,\alpha} \prod_{\jmath\in\mathcal{J}} \left(\frac{v_\jmath - b_{\jmath,x}}{r_{\jmath,x}}\right)^{\alpha_\jmath}.$$

The coefficient sum is unchanged by this relabeling:

$$\sum_{\alpha\in\mathcal{S}_x} |\xi_{x,\alpha}| = \sum_{\beta\in\mathcal{B}_x} |\eta_\beta|.$$

2. Substitute $v_\jmath = b_{\jmath,x} + r_{\jmath,x}z_\jmath$ in the preceding display. Since every $r_{\jmath,x} > 0$,

$$\frac{b_{\jmath,x} + r_{\jmath,x}z_\jmath - b_{\jmath,x}}{r_{\jmath,x}} = z_\jmath.$$

Thus, as a polynomial in $z \in \mathbb{R}^{\mathcal{J}}$,

$$\mathcal{R}_{x,K}(z) = P_{x,K}(b_x + r_x \odot z) - \bar{f}_\epsilon(b_x) = \sum_{\alpha\in\mathcal{S}_x} \xi_{x,\alpha} \prod_{\jmath\in\mathcal{J}} z_\jmath^{\alpha_\jmath}.$$

3. The last identity implies that every coefficient of $\mathcal{R}_{x,K}$ outside $\mathcal{S}_x$ is zero, and that for $\alpha \in \mathcal{A}_x = \operatorname{supp}(\mathcal{R}_{x,K})$ the coefficient $b_{x,\alpha}$ equals $\xi_{x,\alpha}$. Hence

$$\sum_{\alpha \in \mathcal{A}_x} |b_{x,\alpha}| \le \sum_{\alpha \in \mathcal{S}_x} |\xi_{x,\alpha}|.$$

4. By the coefficient-envelope part of Lemma 2,

$$\sum_{\beta \in \mathcal{B}_x} |\eta_\beta| \le 2^{40K+20}(1+\epsilon^{-1}) \sum_{\jmath \in \mathcal{J}} r_{\jmath,x}.$$

   Since $K \ge 2$,

$$40K + 20 \le 60K, \qquad 2^{40K+20} \le 2^{60K} = (2^{60})^K.$$

   Also $(1+\epsilon^{-1}) \sum_{\jmath \in \mathcal{J}} r_{\jmath,x} \ge 0$, because $\epsilon > 0$ and every radius is positive. Combining these inequalities with the relabeling identity gives

$$\sum_{\alpha \in \mathcal{A}_x} |b_{x,\alpha}| \le (2^{60})^K (1+\epsilon^{-1}) \sum_{\jmath \in \mathcal{J}} r_{\jmath,x}.$$

5. It remains to record the coordinate-degree bound. Let $\alpha \in \mathcal{A}_x$. If $\alpha \notin \mathcal{S}_x$, the polynomial identity in Item 2 gives $b_{x,\alpha} = 0$, contradicting $\alpha \in \operatorname{supp}(\mathcal{R}_{x,K})$. Hence $\alpha = \Phi\beta$ for some $\beta \in \mathcal{B}_x$. The normalized-degree part of Lemma 2 gives

$$\beta_\varpi \le 2(K-1) \qquad \text{for every } \varpi \in \{1,2,3,4\}.$$

   Therefore, for each $\jmath \in \mathcal{J}$,

$$\alpha_\jmath = (\Phi\beta)_\jmath = \beta_{\iota_{\mathcal{J}}(\jmath)} \le 2(K-1).$$

□

The centered form in Lemma 7 aligns with the polynomial $U_h(N; z)$ from Algorithm 1. This is the same unbiased polynomial-lifting principle used for nonsmooth functional estimation under Poisson sampling [Jiao et al., 2015].

Definition 21. [`synth_5`] (Pilot-failure contributions) *Fix a cell x, a four-cell mass vector q_x, Poisson scale $m > 0$, and Jackson tuning. Let $\mathbb{P}^{\mathrm{pil,eval}}_{m,q_x}$ denote the product-Poisson law of the pilot and evaluation count vectors, and let G_x be the coordinatewise good-pilot event used for the pilot rectangle in the Jackson–factorial construction. With T_x^{JF} denoting the clipped cell statistic built from the pilot and evaluation counts, define*

$$\mathrm{PF}_{1,x} := \int_{\mathsf{G}_x^c} \left| T_x^{\mathrm{JF}} - \bar{f}_\epsilon(q_x) \right| d\mathbb{P}^{\mathrm{pil,eval}}_{m,q_x}, \qquad \mathrm{PF}_{2,x} := \int_{\mathsf{G}_x^c} \left(T_x^{\mathrm{JF}} - \bar{f}_\epsilon(q_x) \right)^2 d\mathbb{P}^{\mathrm{pil,eval}}_{m,q_x}.$$

The quantities in Definition 21 separate the contribution of pilot failures from the conditional bias and variance calculations on G_x. The final lemma assembles the distributional identities, the conditional factorial moments, and the resulting cellwise controls.

Lemma 8. [`lem:centered-factorial-pilot-control`] (Centered factorial pilot control) *Assume the product experiment satisfies the i.i.d. sampling law in Assumption 1. There is a universal Jackson tuning with self-normalized Poisson pilot-radius constant, polynomial-degree constant $\kappa = 1/100000$, and cutoff $D_0 = 3$. For every overlap parameter $\epsilon \in (0, 1/2)$, there is a constant $C_\epsilon > 0$ such that the following assertions hold for every n, d, every observed law $\mathbb{P}$ on $[d] \times \{0,1\}^2$, and every positive sample size n, whenever $\mathbb{P} \in \mathcal{D}^{\mathrm{obs}}_{d,\epsilon}$ as in Definition 3. Set $m = n/8$, let $q_x = (q_{00,x}, q_{01,x}, q_{10,x}, q_{11,x})$ be the four-coordinate cell vector, and let p_x be the marginal mass of cell x.*

In the uncapped marked experiment generated by

$$M \sim \mathrm{Pois}(n/4)$$

together with i.i.d. observations from $\mathbb{P}$ and independent fair pilot/evaluation marks, the count M has law $\mathrm{Pois}(n/4)$. The induced pilot/evaluation table has law equal to the product, over cells $x \in [d]$ and coordinates $\jmath \in \mathcal{J}$, of two independent Poisson count arrays with coordinate means $mq_{\jmath,x}$. For each fixed cell x, the marginal law of the pilot and evaluation counts in that cell is therefore

$$\bigotimes_{\jmath \in \mathcal{J}} \mathrm{Pois}(mq_{\jmath,x}) \otimes \bigotimes_{\jmath \in \mathcal{J}} \mathrm{Pois}(mq_{\jmath,x}).$$

For every cell x, every fixed pilot table, every coordinate $\jmath \in \mathcal{J}$, every $z \in \mathbb{R}$, and every $h, t \in \mathbb{N}$, conditional expectation over the evaluation counts satisfies

$$\mathbb{E}\big[U_h(N_{\jmath,x}; z) \mid N'\big] = (q_{\jmath,x} - z)^h$$

and

$$\mathbb{E}\big[U_h(N_{\jmath,x}; z) U_t(N_{\jmath,x}; z) \mid N'\big] = \sum_{\ell=0}^{\min\{h,t\}} \binom{h}{\ell}\binom{t}{\ell} \ell! \left(\frac{q_{\jmath,x}}{m}\right)^{\ell} (q_{\jmath,x} - z)^{h+t-2\ell}.$$

With the same tuning, the clipped Jackson-factorial cell statistic T_x^{JF} obeys, uniformly over all cells $x \in [d]$,

$$\big|\mathbb{E}T_x^{\mathrm{JF}} - \bar{f}_\epsilon(q_x)\big| \le C_\epsilon \left\{\sqrt{\frac{p_x}{mL_d}} + \frac{1}{mL_d}\right\},$$

$$\mathrm{Var}(T_x^{\mathrm{JF}}) \le C_\epsilon d^{1/16} \left\{\frac{p_x L_d}{m} + \frac{L_d^2}{m^2}\right\},$$

and the pilot-failure contributions satisfy

$$\mathrm{PF}_{1,x} \le C_\epsilon \left\{\sqrt{\frac{p_x}{mL_d}} + \frac{1}{mL_d}\right\}, \qquad \mathrm{PF}_{2,x} \le C_\epsilon d^{1/16} \left\{\frac{p_x L_d}{m} + \frac{L_d^2}{m^2}\right\}.$$

Proof of Lemma 8.

1. *Tuning and constants.* Use the canonical Jackson tuning specified in the statement, whose numerical components are

$$H_0 = 1024, \qquad \kappa = \frac{1}{100000}, \qquad D_0 = 3.$$

Fix $\epsilon \in (0, 1/2)$ and set

$$F = 1 + \epsilon^{-1}.$$

Let $C_J > 0$ be the constant supplied by Lemma 4, and let $C_1, C_2 > 0$ be the $t = 1$ and $t = 2$ constants in Lemma 5. Define

$$C_\epsilon = 1 + \{5 \cdot 10^{16} C_J + 450000000\, e^{1000} F\} + 450 F C_1 \\ + 8e^{1000} F^2 (2 \cdot 10^6)^2 + 8F^2 C_2.$$

All summands after the leading 1 are nonnegative, so $C_\epsilon > 0$.

2. *Cell notation.* Fix $n, d, \mathbb{P}$, assume $n > 0$ and $\mathbb{P} \in \mathcal{D}^{\text{obs}}_{d,\epsilon}$ as in Definition 3, and put

$$m = \frac{n}{8}, \qquad L = L_d = \log(ed), \qquad \mathcal{J} = \{0,1\} \times \{0,1\}.$$

Then $m > 0$. If $[d]$ is empty, the cellwise assertions are vacuous. Otherwise fix $x \in [d]$. Write

$$q_\jmath = q_{\jmath,x}, \qquad q_x = (q_\jmath)_{\jmath \in \mathcal{J}}, \qquad p_x = \sum_{\jmath \in \mathcal{J}} q_\jmath.$$

The coordinates satisfy $q_\jmath \geq 0$. Since $x \in [d]$, $d \geq 1$, and hence

$$L = \log(ed) = 1 + \log d \geq 1, \qquad L > 0.$$

3. *Poisson splitting.* The uncapped experiment draws $M \sim \text{Pois}(n/4)$, giving the asserted marginal law of M. Conditional on inclusion, an observation in atom $(x', \jmath) \in [d] \times \mathcal{J}$ is independently marked pilot or evaluation with probability $1/2$. Since $n/4 = 2m$, Poisson thinning gives pilot and evaluation intensities

$$2m\, q_{\jmath,x'} \cdot \frac{1}{2} = m q_{\jmath,x'}.$$

Thus

$$\text{Law}\{(N'_{\jmath,x'}, N_{\jmath,x'})_{x',\jmath}\} = \bigotimes_{x' \in [d]} \bigotimes_{\jmath \in \mathcal{J}} \text{Pois}(m q_{\jmath,x'}) \otimes \bigotimes_{x' \in [d]} \bigotimes_{\jmath \in \mathcal{J}} \text{Pois}(m q_{\jmath,x'}).$$

Projecting this product law to the fixed cell x gives

$$\bigotimes_{\jmath \in \mathcal{J}} \text{Pois}(m q_\jmath) \otimes \bigotimes_{\jmath \in \mathcal{J}} \text{Pois}(m q_\jmath).$$

4. *Centered factorial moments.* By Algorithm 1,

$$U_h(N; z) = \sum_{a=0}^{h} \binom{h}{a} (-z)^{h-a} \frac{(N)_a}{m^a}.$$

Equivalently,

$$\sum_{h=0}^{\infty} U_h(N; z) \frac{w^h}{h!} = e^{-zw} \left(1 + \frac{w}{m}\right)^N.$$

For $N \sim \text{Pois}(mq)$,

$$E\left[e^{-zw} \left(1 + \frac{w}{m}\right)^N\right] = e^{(q-z)w},$$

so coefficient comparison gives $E\,U_h(N;z) = (q-z)^h$. With two formal variables,

$$E\left[e^{-z(w_1+w_2)}\left(1+\frac{w_1}{m}\right)^N\left(1+\frac{w_2}{m}\right)^N\right] = \exp\left\{(q-z)(w_1+w_2)+\frac{q}{m}w_1w_2\right\}.$$

The coefficient of $w_1^h w_2^t/(h!t!)$ is therefore

$$\sum_{\ell=0}^{\min\{h,t\}}\binom{h}{\ell}\binom{t}{\ell}\ell!\left(\frac{q}{m}\right)^{\ell}(q-z)^{h+t-2\ell}.$$

The evaluation counts are independent of the pilot counts, so with $q = q_{\jmath,x}$ both identities hold conditionally on every fixed pilot table.

5. *Pilot geometry and failure terms.* For a fixed pilot table, define

$$c_{\jmath,x} = \frac{N'_{\jmath,x}}{m}, \qquad h_{\jmath,x} = 1024\left\{\sqrt{\frac{c_{\jmath,x}L}{m}}+\frac{L}{m}\right\},$$

$$\ell_{\jmath,x} = \max\{0, c_{\jmath,x}-h_{\jmath,x}\}, \qquad u_{\jmath,x} = c_{\jmath,x}+h_{\jmath,x},$$

and

$$b_{\jmath,x} = \frac{\ell_{\jmath,x}+u_{\jmath,x}}{2}, \qquad r_{\jmath,x} = \frac{u_{\jmath,x}-\ell_{\jmath,x}}{2}.$$

Let

$$Q_x = \prod_{\jmath\in\mathcal{J}}[\ell_{\jmath,x}, u_{\jmath,x}], \qquad S_x = F\sum_{\jmath\in\mathcal{J}} r_{\jmath,x}, \qquad \varrho_x = d^{1/4}S_x,$$

and define the good-pilot event

$$\mathsf{G}_x = \bigcap_{\jmath\in\mathcal{J}}\left\{|c_{\jmath,x}-q_\jmath| \le \frac{1}{4}h_{\jmath,x}\right\}.$$

On G_x, $q_x \in Q_x$: if $c_{\jmath,x}-h_{\jmath,x} \le 0$, then $\ell_{\jmath,x} = 0 \le q_\jmath$ and $q_\jmath \le c_{\jmath,x}+h_{\jmath,x}/4 \le u_{\jmath,x}$; if $c_{\jmath,x}-h_{\jmath,x} > 0$, then $c_{\jmath,x}-h_{\jmath,x} \le q_\jmath \le c_{\jmath,x}+h_{\jmath,x}$. Also $r_{\jmath,x} > 0$ and

$$\frac{1}{2}h_{\jmath,x} \le r_{\jmath,x} \le h_{\jmath,x}.$$

Define

$$\Gamma_x(N') = \sum_{\jmath\in\mathcal{J}}\left[|c_{\jmath,x}-q_\jmath| + 1024\left\{\sqrt{\frac{c_{\jmath,x}L}{m}}+\frac{L}{m}\right\}\right].$$

The interval definitions give

$$\|b_x-q_x\|_1 + \sum_{\jmath\in\mathcal{J}} r_{\jmath,x} \le 2\Gamma_x(N').$$

Using the notation of Definitions 7 and 8, for any two nonnegative four-vectors u, v, the global cell extension satisfies the four-coordinate Lipschitz bound

$$|\bar{f}_\epsilon(u)-\bar{f}_\epsilon(v)| \le F\sum_{\jmath\in\mathcal{J}}|u_\jmath - v_\jmath|.$$

For a fixed arm a, write

$$\rho_a(w) = \max\{s_a(w), \epsilon s(w)\},$$

so $g_{a,\epsilon}(w) = 0$ when $s(w) = 0$, and $g_{a,\epsilon}(w) = s(w)w_{a1}/\rho_a(w)$ otherwise. If one of $s(u), s(v)$ is zero, all coordinates of that vector vanish and $0 \le g_{a,\epsilon}(w) \le s(w)$ gives the bound. On the positive-mass part, split according to whether $s_a(w) \le \epsilon s(w)$ or $s_a(w) \ge \epsilon s(w)$ for $w = u, v$. In the threshold-threshold case the arm value is w_{a1}/ϵ. In the arm-active case use

$$\left|\frac{u_{a1}}{s_a(u)} - \frac{v_{a1}}{s_a(v)}\right| \le \frac{|u_{a1} - v_{a1}| + |u_{a0} - v_{a0}|}{s_a(u)}$$

and $s(u)/s_a(u) \le \epsilon^{-1}$. In the two mixed cases, the nonnegative gap between $s_a(w)$ and $\epsilon s(w)$ is bounded by the corresponding changes in s_a and s, hence by $\sum_\jmath |u_\jmath - v_\jmath|$; the preceding estimates give the same F-Lipschitz bound for each $g_{a,\epsilon}$. Taking the maximum over the two arms preserves this bound. The clipping interval in Algorithm 1 therefore gives

$$|T_x^{\mathrm{JF}} - \bar{f}_\epsilon(q_x)| \le F\|b_x - q_x\|_1 + d^{1/4} F \sum_{\jmath \in \mathcal{J}} r_{\jmath,x} \le 2d^{1/4} F\,\Gamma_x(N').$$

With $\mu_x^{\mathrm{pil}} = \bigotimes_\jmath \mathrm{Pois}(mq_\jmath)$ and $\mu_x^{\mathrm{eval}} = \bigotimes_\jmath \mathrm{Pois}(mq_\jmath)$, set

$$\mathrm{PF}_{1,x} = \int_{\mathsf{G}_x^c} |T_x^{\mathrm{JF}} - \bar{f}_\epsilon(q_x)|\, d(\mu_x^{\mathrm{pil}} \otimes \mu_x^{\mathrm{eval}})$$

and

$$\mathrm{PF}_{2,x} = \int_{\mathsf{G}_x^c} (T_x^{\mathrm{JF}} - \bar{f}_\epsilon(q_x))^2\, d(\mu_x^{\mathrm{pil}} \otimes \mu_x^{\mathrm{eval}}).$$

6. *Good-pilot scales.* By Lemma 6, on G_x,

$$\sum_{\jmath \in \mathcal{J}} r_{\jmath,x} \le 2 \cdot 10^6 \left\{\sqrt{\frac{p_x L}{m}} + \frac{L}{m}\right\}.$$

Set

$$v_x = \sqrt{\frac{p_x L}{m}} + \frac{L}{m}, \qquad \eta_x = \sqrt{\frac{p_x}{mL}} + \frac{1}{mL}.$$

Since $L \ge 1$,

$$v_x = L\sqrt{\frac{p_x}{mL}} + \frac{L}{m} \le L^2 \eta_x.$$

For $K_d = \max\{2, \lfloor L/100000 \rfloor\}$, first

$$\frac{L}{K_d} \le 150000, \qquad \left(\frac{L}{K_d}\right)^2 \le 150000^2.$$

Indeed, if $L \le 3 \cdot 10^5$, then $K_d \ge 2$ gives $L/K_d \le 150000$. If $L > 3 \cdot 10^5$, then

$$K_d \ge \frac{L}{100000} - 1 \ge \frac{L}{150000},$$

and the same bound follows.

The logarithmic calibrations used below are

$$Ld^{-3/16} \le 7, \qquad Ld^{1/4}e^{-20L} \le 7, \qquad d^{1/2}e^{-20L} \le d^{1/16},$$

and

$$L^2 d^{-3/16} \le 225, \qquad L^2 d^{1/4} e^{-20L} \le 225.$$

For $d \ge 1$, the calculus inequality $\log y \le y^a/a$ for $y \ge 1$ gives, with $a = 3/16$,

$$L = 1 + \log d \le d^{3/16} + \frac{16}{3} d^{3/16} \le 7d^{3/16},$$

which proves $Ld^{-3/16} \le 7$. Since $L \ge 1$,

$$e^{-20L} \le e^{-L} = \frac{1}{ed} \le \frac{1}{d}.$$

Consequently

$$Ld^{1/4}e^{-20L} \le Ld^{-3/4} \le Ld^{-3/16} \le 7,$$

and

$$d^{1/2}e^{-20L} \le d^{-1/2} \le d^{1/16}.$$

With $a = 3/32$, the same calculus inequality gives

$$L = 1 + \log d \le d^{3/32} + \frac{32}{3} d^{3/32} \le 15d^{3/32},$$

so

$$L^2 \le 225 d^{3/16}.$$

Multiplying by $d^{-3/16}$ proves $L^2 d^{-3/16} \le 225$, and using $e^{-20L} \le 1/d$ once more gives

$$L^2 d^{1/4} e^{-20L} \le L^2 d^{-3/4} \le L^2 d^{-3/16} \le 225.$$

7. *Good-pilot factorial L_2 bound.* Fix a pilot table in G_x, and write

$$\chi_{\jmath,x} = \sqrt{\frac{c_{\jmath,x} L}{m}}, \qquad h_{\jmath,x} = 1024 \left(\chi_{\jmath,x} + \frac{L}{m} \right).$$

First,

$$\frac{q_\jmath}{m r_{\jmath,x}^2} \le \frac{1}{L}.$$

Indeed, the good-pilot inequality gives

$$q_\jmath \le c_{\jmath,x} + \frac{1}{4} h_{\jmath,x} \le 384 \left(c_{\jmath,x} + \frac{L}{m} \right),$$

where the last inequality uses $\chi_{\jmath,x} \le (c_{\jmath,x} + L/m)/2$. Since $r_{\jmath,x} \ge h_{\jmath,x}/2 = 512(\chi_{\jmath,x} + L/m)$,

$$m r_{\jmath,x}^2 \ge m\, 512^2 \left(\chi_{\jmath,x} + \frac{L}{m} \right)^2 \ge 384 L \left(c_{\jmath,x} + \frac{L}{m} \right),$$

because $\chi_{\jmath,x}^2 = c_{\jmath,x} L/m$. Combining the last two displays proves the ratio bound.

Also $q_x \in Q_x$ gives

$$|q_\jmath - b_{\jmath,x}| \le r_{\jmath,x}.$$

For any nonnegative integer h, the centered-factorial second-moment identity from Item 4 yields

$$E\left[\left\{\frac{U_h(N_{\jmath,x}; b_{\jmath,x})}{r_{\jmath,x}^h}\right\}^2 \middle| N'\right] \le \sum_{\ell=0}^{h} \binom{h}{\ell}^2 \ell! L^{-\ell} \le \exp(h^2/L).$$

Let P_{x,K_d} be the Jackson tensor polynomial on Q_x, and write the centered normalized expansion from Lemma 7 as

$$P_{x,K_d}(v) - \bar{f}_\epsilon(b_x) = \sum_{\alpha \in \mathcal{A}_x} b_{x,\alpha} \prod_{\jmath \in \mathcal{J}} \left(\frac{v_\jmath - b_{\jmath,x}}{r_{\jmath,x}}\right)^{\alpha_\jmath},$$

with

$$\sum_{\alpha \in \mathcal{A}_x} |b_{x,\alpha}| \le (2^{60})^{K_d} F \sum_{\jmath \in \mathcal{J}} r_{\jmath,x}, \qquad \alpha_\jmath \le 2(K_d - 1).$$

Define the factorial lift

$$\mathcal{L}_x(N) = \sum_{\alpha \in \mathcal{A}_x} b_{x,\alpha} \prod_{\jmath \in \mathcal{J}} \frac{U_{\alpha_\jmath}(N_{\jmath,x}; b_{\jmath,x})}{r_{\jmath,x}^{\alpha_\jmath}}.$$

Independence of the evaluation coordinates and Minkowski's inequality give

$$E\{\mathcal{L}_x(N)^2 \mid N'\} \le (2^{60})^{2K_d} F^2 \left(\sum_{\jmath \in \mathcal{J}} r_{\jmath,x}\right)^2 \exp(16K_d^2/L).$$

The calibrated coefficient growth bound is

$$(2^{60})^{2K_d} \exp(16K_d^2/L) \le e^{1000} d^{1/16}.$$

Indeed, $K_d = \max\{2, \lfloor L/100000 \rfloor\}$ and the floor bound give

$$K_d \le 2 + \frac{L}{100000}.$$

Since $L \ge 1$, squaring the last display gives

$$K_d^2 \le L\left(5 + \frac{L}{10^{10}}\right), \qquad \frac{K_d^2}{L} \le 5 + \frac{L}{10^{10}}.$$

Also $L = 1 + \log d$, and hence

$$120K_d + 16\frac{K_d^2}{L} \le 120\left(2 + \frac{L}{100000}\right) + 16\left(5 + \frac{L}{10^{10}}\right) \le 1000 + \frac{\log d}{16}.$$

Because $2 < e$,

$$(2^{60})^{2K_d} = 2^{120K_d} \le e^{120K_d}.$$

Multiplying by the remaining exponential factor and using the preceding exponent comparison yields

$$(2^{60})^{2K_d} \exp(16K_d^2/L) \le \exp\left(120K_d + 16\frac{K_d^2}{L}\right) \le \exp\left(1000 + \frac{\log d}{16}\right) = e^{1000} d^{1/16}.$$

Therefore

$$E\{\mathcal{L}_x(N)^2 \mid N'\} \le e^{1000} d^{1/16} S_x^2.$$

8. *Good-pilot bias.* For the fixed good pilot, the factorial identities give

$$E\{\mathcal{L}_x(N) \mid N'\} = P_{x,K_d}(q_x) - \bar{f}_\epsilon(b_x).$$

Let

$$Z_x = \bar{f}_\epsilon(b_x) + \mathcal{L}_x(N), \qquad T_x^{\mathrm{JF}} = \Pi_{[\bar{f}_\epsilon(b_x)-\varrho_x,\, \bar{f}_\epsilon(b_x)+\varrho_x]}(Z_x).$$

Hence

$$E(T_x^{\mathrm{JF}} \mid N') - \bar{f}_\epsilon(q_x) = \{P_{x,K_d}(q_x) - \bar{f}_\epsilon(q_x)\} + E(T_x^{\mathrm{JF}} - Z_x \mid N').$$

For each coordinate,

$$\sqrt{(q_\jmath - \ell_{\jmath,x})(u_{\jmath,x} - q_\jmath)} \le 10^7 \sqrt{\frac{q_\jmath L}{m}}.$$

To see this, if $c_{\jmath,x} - h_{\jmath,x} \le 0$, then $\ell_{\jmath,x} = 0$, and the quadratic inequality $c_{\jmath,x} \le h_{\jmath,x}$ gives $c_{\jmath,x} \le 1.1 \cdot 10^6 L/m$ and $h_{\jmath,x} \le 1.2 \cdot 10^9 L/m$, so

$$(q_\jmath - \ell_{\jmath,x})(u_{\jmath,x} - q_\jmath) \le q_\jmath(c_{\jmath,x} + h_{\jmath,x}) \le 10^{14} \frac{q_\jmath L}{m}.$$

If $c_{\jmath,x} - h_{\jmath,x} > 0$, then the good-pilot inequality gives $q_\jmath \ge 3c_{\jmath,x}/4$ and $L/m \le c_{\jmath,x}$, whence

$$h_{\jmath,x} \le 4096\sqrt{\frac{q_\jmath L}{m}}, \qquad \sqrt{(q_\jmath - \ell_{\jmath,x})(u_{\jmath,x} - q_\jmath)} \le r_{\jmath,x} = h_{\jmath,x}.$$

Applying Lemma 4 at $q_x \in Q_x$ gives

$$|P_{x,K_d}(q_x) - \bar{f}_\epsilon(q_x)| \le C_J \sum_{\jmath \in \mathcal{J}} \left\{ \frac{10^7\sqrt{q_\jmath L/m}}{K_d} + \frac{r_{\jmath,x}}{K_d^2} \right\}.$$

Since $\sum_\jmath \sqrt{q_\jmath L/m} \le 2\sqrt{p_x L/m}$ and $\sum_\jmath r_{\jmath,x} \le 2 \cdot 10^6 v_x$,

$$\sum_{\jmath \in \mathcal{J}} \left\{ \frac{10^7\sqrt{q_\jmath L/m}}{K_d} + \frac{r_{\jmath,x}}{K_d^2} \right\} \le 22 \cdot 10^6 \frac{\sqrt{p_x L/m}}{K_d} + 2 \cdot 10^6 \frac{L/m}{K_d^2} \le 5 \cdot 10^{16} \eta_x.$$

From the interval projection definition in Definition 1, a direct case split gives the following elementary consequence: for every real midpoint mid, every real input raw, and every radius rad > 0,

$$\left|\Pi_{[\mathrm{mid}-\mathrm{rad},\mathrm{mid}+\mathrm{rad}]}(\mathrm{raw}) - \mathrm{raw}\right| \le \frac{(\mathrm{raw}-\mathrm{mid})^2}{\mathrm{rad}}.$$

If raw lies inside the interval, the left side is zero. If raw $<$ mid $-$ rad, then the inequality follows after multiplying by rad from

$$(\mathrm{raw}-\mathrm{mid})^2 + \mathrm{rad}(\mathrm{raw}-\mathrm{mid}) + \mathrm{rad}^2 \ge 0,$$

and the case raw $>$ mid $+$ rad is identical with the middle sign reversed. Applying this with midpoint $\bar{f}_\epsilon(b_x)$, radius ϱ_x, and input Z_x, and using $Z_x - \bar{f}_\epsilon(b_x) = \mathcal{L}_x(N)$, gives

$$|T_x^{\mathrm{JF}} - Z_x| \le \frac{\mathcal{L}_x(N)^2}{\varrho_x}.$$

Using the factorial L_2 bound from Item 7,

$$|E(T_x^{\mathrm{JF}} - Z_x \mid N')| \le \frac{e^{1000} d^{1/16} S_x^2}{d^{1/4} S_x} = e^{1000} F d^{-3/16} \sum_{\jmath \in \mathcal{J}} r_{\jmath,x}.$$

Together with the good-pilot radius bound and $L^2 d^{-3/16} \le 225$, this implies

$$\left|E(T_x^{\mathrm{JF}} \mid N') - \bar{f}_\epsilon(q_x)\right| \le \{5 \cdot 10^{16} C_J + 450000000\, e^{1000} F\} \eta_x \qquad \text{on } \mathsf{G}_x.$$

9. *Good-pilot second moment.* On G_x, the four-coordinate Lipschitz bound from Item 5 and $q_x \in Q_x$ give

$$|\bar{f}_\epsilon(b_x) - \bar{f}_\epsilon(q_x)| \le S_x.$$

A direct case split using the interval projection $\Pi_{[a,b]}$ from Definition 1 gives the contraction

$$\left|\Pi_{[\mathrm{mid}\,-\,\mathrm{rad},\mathrm{mid}\,+\,\mathrm{rad}]}(\mathrm{raw}) - \mathrm{mid}\right| \le |\,\mathrm{raw} - \mathrm{mid}\,|.$$

Applying it with midpoint $\bar{f}_\epsilon(b_x)$, radius ϱ_x, and input Z_x gives

$$|T_x^{\mathrm{JF}} - \bar{f}_\epsilon(b_x)| \le |\mathcal{L}_x(N)|.$$

Therefore

$$|T_x^{\mathrm{JF}} - \bar{f}_\epsilon(q_x)| \le |\mathcal{L}_x(N)| + S_x,$$

and hence

$$(T_x^{\mathrm{JF}} - \bar{f}_\epsilon(q_x))^2 \le 2\mathcal{L}_x(N)^2 + 2S_x^2.$$

After conditional expectation over the evaluation counts,

$$E\left[\{T_x^{\mathrm{JF}} - \bar{f}_\epsilon(q_x)\}^2 \mid N'\right] \le 2e^{1000} d^{1/16} S_x^2 + 2S_x^2 \le 4e^{1000} d^{1/16} S_x^2,$$

because $1 \le e^{1000} d^{1/16}$. Integrating over pilots on G_x and using $\sum_\jmath r_{\jmath,x} \le 2 \cdot 10^6 v_x$ gives

$$E\left[\{T_x^{\mathrm{JF}} - \bar{f}_\epsilon(q_x)\}^2 \mathbf{1}_{\mathsf{G}_x}\right] \le 4e^{1000} d^{1/16} (2 \cdot 10^6 F v_x)^2.$$

10. *Bad-pilot bounds.* The pointwise domination from Item 5 and Lemma 5 give

$$\mathrm{PF}_{1,x} \le 2d^{1/4} F C_1 e^{-20L} v_x$$

and

$$\mathrm{PF}_{2,x} \le (2d^{1/4} F)^2 C_2 e^{-20L} v_x^2.$$

Using $v_x \le L^2 \eta_x$ and $L^2 d^{1/4} e^{-20L} \le 225$,

$$\mathrm{PF}_{1,x} \le 450 F C_1\, \eta_x.$$

Also

$$v_x^2 = \left(\sqrt{\frac{p_x L}{m}} + \frac{L}{m}\right)^2 \le 2\left\{\frac{p_x L}{m} + \frac{L^2}{m^2}\right\},$$

and $d^{1/2} e^{-20L} \le d^{1/16}$, so

$$\mathrm{PF}_{2,x} \le 8F^2 C_2 d^{1/16} \left\{\frac{p_x L}{m} + \frac{L^2}{m^2}\right\}.$$

The displayed estimates are exactly the asserted pilot-failure bounds after enlarging by C_ϵ.

11. *Bias.* Disintegrating over the pilot table, the good-pilot conditional bound from Item 8 and the bad-pilot contribution give

$$\left|ET_x^{\mathrm{JF}} - \bar{f}_\epsilon(q_x)\right| \le \{5 \cdot 10^{16} C_J + 450000000\, e^{1000} F\}\eta_x + \mathrm{PF}_{1,x}.$$

Using the estimate for $\mathrm{PF}_{1,x}$ and the definition of C_ϵ,

$$\left|ET_x^{\mathrm{JF}} - \bar{f}_\epsilon(q_x)\right| \le C_\epsilon \left\{\sqrt{\frac{p_x}{mL_d}} + \frac{1}{mL_d}\right\}.$$

12. *Variance.* Combining the good-pilot second-moment estimate with $\mathrm{PF}_{2,x}$ yields

$$E\{T_x^{\mathrm{JF}} - \bar{f}_\epsilon(q_x)\}^2 \le 4e^{1000} d^{1/16}(2 \cdot 10^6 F v_x)^2 + \mathrm{PF}_{2,x}.$$

Using the bound on v_x^2, the estimate for $\mathrm{PF}_{2,x}$, and the definition of C_ϵ,

$$E\{T_x^{\mathrm{JF}} - \bar{f}_\epsilon(q_x)\}^2 \le C_\epsilon d^{1/16} \left\{\frac{p_x L_d}{m} + \frac{L_d^2}{m^2}\right\}.$$

Variance is bounded by the mean squared deviation about any fixed center, so

$$\mathrm{Var}(T_x^{\mathrm{JF}}) \le E\{T_x^{\mathrm{JF}} - \bar{f}_\epsilon(q_x)\}^2.$$

Thus the stated bias, variance, and pilot-failure estimates hold uniformly in the cell x, in n, d, and in $\mathbb{P} \in \mathcal{D}_{d,\epsilon}^{\mathrm{obs}}$.

□

Lemma 8 supplies the cellwise estimates used to prove the upper bound. The first displayed identities identify the Poissonized marked experiment with independent pilot and evaluation arrays. The factorial identities then center the polynomial at the pilot rectangle center, and the final bounds control bias, variance, and pilot-failure terms at the same local scale.

B Moment matching and lower-bound details

This appendix gives the dense lower-bound construction used in Theorem 2. The argument works inside an equal-mass, equal-propensity submodel and converts best polynomial approximation of $|t|$ into a separation in the optimal-value target. Its testing step follows the classical moment-matching method for nonsmooth functionals [Cai and Low, 2011], with Poissonized mixtures used to make the likelihood calculation cellwise.

We first state the Poissonized risk that appears in the comparison. It is the same minimax squared loss criterion as Definition 5, evaluated under a Poisson observed-sample law and with the estimator projected to the target range.

Definition 22. [`synth_16`] (Poisson minimax risk) *For $m, d \in \mathbb{N}$ and $\epsilon \in \mathbb{R}$, define the Poissonized minimax squared risk*

$$\mathfrak{R}_{m,d,\epsilon}^{\mathrm{Pois}} := \inf_{\widehat{V}} \sup_{\mathbb{P} \in \mathcal{D}_{d,\epsilon}^{\mathrm{obs}}} \mathbb{E}_{\mathrm{Pois}(m,\mathbb{P})} \left[\left\{\Pi_{[0,1]}\big(\widehat{V}(\mathcal{S})\big) - \Psi(\mathbb{P})\right\}^2\right].$$

Here $\mathcal{S}$ is the finite observed sample drawn from the Poisson observed-sample law with mean m and atom law $\mathbb{P}$, the supremum ranges over the observed-law model class in Definition 3, the target $\Psi(\mathbb{P})$ is the observed optimal value from Definition 4, and the infimum ranges over all dense Poisson estimators on the d-cell observed sample space.

The submodel is indexed by a dense contrast vector. Each cell has mass $1/d$, treatment probability $1/2$, and opposite perturbations of the two arm-specific outcome means.

Definition 23. [`synth_15`] (Dense contrast law) *For an integer $d \geq 2$ and a contrast vector $\theta = (\theta_x)_{x\in[d]}$ with $\theta_x \in [-1/2, 1/2]$ for every $x \in [d]$, define $P_\theta^{\mathrm{dense}}$ as the full-data probability law on $(X, A, Y, Y(0), Y(1))$ whose atom at $(x, a, y, y_0, y_1) \in [d] \times \{0,1\}^4$ is*

$$P_\theta^{\mathrm{dense}}(x, a, y, y_0, y_1) = \begin{cases} \dfrac{1}{2d}\, b_{(1-\theta_x)/2}(y_0)\, b_{(1+\theta_x)/2}(y_1), & y = y_a, \\ 0, & y \neq y_a, \end{cases}$$

where $y_a = y_1$ if $a = 1$ and $y_a = y_0$ if $a = 0$, and $b_\mu(1) = \mu$, $b_\mu(0) = 1 - \mu$.

For this law, the target in Definition 4 becomes an average of $|\theta_x|$ plus the common baseline. The lower bound therefore reduces estimation of the value to estimation of an absolute-value functional over many small coordinates.

The construction below fixes the moment-matching degree, scales the contrast amplitude, and defines the two product priors. The existence of the symmetric measures $\nu_{0,K}$ and $\nu_{1,K}$ with matching moments through degree K and separation $2E_K$ is the approximation-theoretic input from Cai and Low [2011, Lemma 1 and Section 3.1]; the same source gives constants $0 < c < C < \infty$ with $c/K \leq E_K \leq C/K$.

Definition 24. [`def:dense-moment-matching-construction`] (Dense moment-matching construction) *For $n, d \in \mathbb{N}$, $\epsilon \in \mathbb{R}$, the overlap parameter, a dense domain certificate*

$$2 \leq d, \qquad 0 < \epsilon, \qquad \epsilon < \frac{1}{2}, \qquad 0 < a_{n,d} \leq \frac{1}{2},$$

and a dense prior family $K \mapsto (\nu_{0,K}, \nu_{1,K})$ indexed by positive even degrees, define the dense moment-matching construction as follows. For every positive even K, the family carries the prior-pair conditions that $\nu_{0,K}$ and $\nu_{1,K}$ are probability measures supported on $[-1, 1]$, are symmetric, have matching moments through degree K, and satisfy

$$\int_{-1}^{1} |t|\, d\nu_{1,K}(t) - \int_{-1}^{1} |t|\, d\nu_{0,K}(t) = 2E_K.$$

The construction has components:

- ***(Dense law.)*** *For each $\theta \in [-1/2, 1/2]^d$, $P_\theta^{\mathrm{dense}}$ is the potential-outcome law induced by the dense full-data masses, equivalently $p_x = 1/d$, $\pi_x = 1/2$, $\mu_{1x} = (1 + \theta_x)/2$, and $\mu_{0x} = (1 - \theta_x)/2$, with $Y(0)$ and $Y(1)$ conditionally independent Bernoulli variables given $X = x$, $A \mid X = x \sim \mathrm{Bernoulli}(1/2)$ independently of the potential outcomes, and $Y = Y(A)$.*

- ***(Approximation scale.)*** *The approximation degree and best even-degree approximation error are*

$$K_d^{\mathrm{lb}} = 2\lceil 4L_d \rceil, \qquad E_K = \inf\left\{ e \in \mathbb{R} : \exists p,\ \deg(p) \leq K,\ \sup_{t\in[-1,1]} \big||t| - p(t)\big| \leq e \right\}.$$

  The construction uses $E_{K_d^{\mathrm{lb}}}$.

- ***(Poisson intensity and amplitude.)*** *The one-cell Poisson intensity and dense contrast amplitude are*

$$\lambda_{n,d} = \frac{2n}{d}, \qquad a_{n,d} = \sqrt{\frac{K_d^{\mathrm{lb}}}{64\lambda_{n,d}}}.$$

- ***(Coordinate priors.)*** *The coordinate priors are*

$$\nu_{0,K_d^{\mathrm{lb}}}, \qquad \nu_{1,K_d^{\mathrm{lb}}}.$$

- ***(Dense product priors.)*** *The priors $\Pi_{0,n,d}$ and $\Pi_{1,n,d}$ are the pushforwards into the dense contrast class of the restrictions of $\nu_{0,K_d^{\mathrm{lb}}}^{\otimes d}$ and $\nu_{1,K_d^{\mathrm{lb}}}^{\otimes d}$ to $[-1,1]^d$, under coordinatewise scaling by $a_{n,d}$.*

- ***(Observation mixtures.)*** *The mixtures $\mathbb{M}_{0,n,d}$ and $\mathbb{M}_{1,n,d}$ are the observation-kernel mixtures induced by $\Pi_{0,n,d}$ and $\Pi_{1,n,d}$, with independent Poissonized sample size* $\mathrm{Pois}(2n)$*.*

- ***(Risk and separation.)*** *The risk component is the Poissonized minimax risk*

$$\mathfrak{R}^{\mathrm{Pois}}_{2n,d,\epsilon}.$$

  The prior target-mean separation is

$$\Delta^{\mathrm{dense}}_{n,d} = E_{\Pi_{1,n,d}}\Psi\{\mathrm{Obs}(P_\theta^{\mathrm{dense}})\} - E_{\Pi_{0,n,d}}\Psi\{\mathrm{Obs}(P_\theta^{\mathrm{dense}})\}.$$

The construction carries the supplied prior-pair certificate at degree K_d^{lb} and the identities

$$K = K_d^{\mathrm{lb}}, \qquad \Delta^{\mathrm{dense}}_{n,d} = a_{n,d} E_{K_d^{\mathrm{lb}}}.$$

The amplitude condition keeps the constructed contrasts inside the dense law in Definition 23. In the range $d^2 < n$, the logarithmic degree is small enough relative to the one-cell Poisson intensity to ensure this admissibility.

Lemma 9. [`lem:dense-amplitude-range`] (Dense amplitude range) *Let $n, d \in \mathbb{N}$ satisfy $2 \le d$ and $d^2 < n$. For the lower-bound contrast amplitude $a_{n,d}$ defined in Definition 24,*

$$0 < a_{n,d} \le \frac{1}{2}.$$

Proof of Lemma 9. Write

$$L_d = \log(ed), \qquad K_d^{\mathrm{lb}} = 2\lceil 4L_d \rceil, \qquad \lambda_{n,d} = \frac{2n}{d}, \qquad a_{n,d} = \sqrt{\frac{K_d^{\mathrm{lb}}}{64\lambda_{n,d}}},$$

as in Definition 24. Since $2 \le d$, we have $0 < d$ and hence, as real numbers,

$$\log d \le d - 1.$$

Moreover

$$L_d = \log(ed) = 1 + \log d,$$

so

$$L_d \leq d.$$

Also $ed > 1$, and therefore

$$0 < L_d.$$

It follows that

$$0 < \lceil 4L_d \rceil, \qquad 0 < K_d^{\mathrm{lb}},$$

and, using $L_d \leq d$,

$$\lceil 4L_d \rceil \leq 4d, \qquad K_d^{\mathrm{lb}} = 2\lceil 4L_d \rceil \leq 8d.$$

The assumption $d^2 < n$ gives $0 < n$, so

$$0 < \lambda_{n,d} = \frac{2n}{d}.$$

Consequently

$$0 < \frac{K_d^{\mathrm{lb}}}{64\lambda_{n,d}},$$

and hence

$$0 < a_{n,d}.$$

It remains to bound the square-root argument by $1/4$. Since

$$K_d^{\mathrm{lb}} \leq 8d$$

and $d^2 < n$, we have

$$K_d^{\mathrm{lb}} d \leq 8d^2 < 8n \leq 32n.$$

Dividing by the positive quantity $32n$ gives

$$\frac{K_d^{\mathrm{lb}} d}{32n} \leq 1.$$

Equivalently, because $\lambda_{n,d} = 2n/d$,

$$\frac{K_d^{\mathrm{lb}}}{64\lambda_{n,d}} = \frac{K_d^{\mathrm{lb}} d}{128n} \leq \frac{1}{4} = \left(\frac{1}{2}\right)^2.$$

Taking square roots yields

$$a_{n,d} = \sqrt{\frac{K_d^{\mathrm{lb}}}{64\lambda_{n,d}}} \leq \frac{1}{2}.$$

Combining this with $0 < a_{n,d}$ proves the claim. □

The testing comparison is expressed in total variation. The next definitions isolate the distance, the likelihood tail left after moment matching, and the one-cell likelihood ratio used in the product calculation.

Definition 25. [`synth_14`] (Total variation distance) *For probability laws $\mathbb{M}_0$ and $\mathbb{M}_1$ on a common measurable space, define $\mathrm{TV}(\mathbb{M}_0, \mathbb{M}_1) = \sup_B |\mathbb{M}_0(B) - \mathbb{M}_1(B)|$, where the supremum ranges over measurable events B.*

Definition 26. [synth_11] (Dense likelihood tail) *For the dense lower-bound degree K_d^{lb}, Poissonized cell intensity $\lambda_{n,d} = 2n/d$, and dense amplitude $a_{n,d}$, define the dense likelihood tail by*

$$\mathrm{tail}_{n,d}^{\mathrm{dense}} := \sum_{r=K_d^{\mathrm{lb}}+1}^{\infty} \frac{(\lambda_{n,d} a_{n,d}^2)^r}{r!}.$$

Definition 27. [synth_9] (One-cell dense likelihood ratio) *Fix a Poisson cell intensity $\lambda \geq 0$. Let R_+ and R_- be the two sign counts in one dense cell, and let E_0 denote expectation under the zero-contrast baseline law $R_+, R_- \overset{\mathrm{ind}}{\sim} \mathrm{Pois}(\lambda/2)$. For $\theta \in [-1, 1]$, the one-cell likelihood ratio relative to this baseline is*

$$L_\theta(R_+, R_-) := (1+\theta)^{R_+}(1-\theta)^{R_-}.$$

In the dense lower bound, L_{θ_x} applies this definition with the cell contrast θ_x and $\lambda = \lambda_{n,d}$.

Moment matching removes the first K_d^{lb} likelihood terms, so the remaining total-variation control is governed by the tail in Definition 26. The lower bound then applies the standard fuzzy-hypothesis testing reduction of Le Cam [1986] and Tsybakov [2009]: two priors with separated target means, concentrated targets, and small induced total variation force squared risk at the square of the separation.

Lemma 10. [lem:dense-moment-matching-lower] (Dense moment matching lower bound)[5] *Suppose the product experiment satisfies the i.i.d. sampling law in Assumption 1. There is an integer cutoff $D_0 \geq 2$ such that, for every overlap constant $\epsilon \in (0, 1/2)$, there is a constant $c_\epsilon > 0$ with the following property. For all integers d, n satisfying $d \geq D_0$ and $d^2 < n$:*

- ***(Degree and amplitude.)*** *The lower-bound degree and amplitude obey*

$$8L_d \leq K_d^{\mathrm{lb}} < 8L_d + 2, \qquad a_{n,d}^2 = \frac{K_d^{\mathrm{lb}} d}{128n}, \qquad a_{n,d} \leq \frac{1}{2}.$$

- ***(Dense observed submodel.)*** *For every $\theta \in [-a_{n,d}, a_{n,d}]^d$, the dense equal-mass contrast submodel $P_\theta^{\mathrm{dense}}$ has observed margin in $\mathcal{D}_{d,\epsilon}^{\mathrm{obs}}$ and satisfies, for every $x \in [d]$,*

$$p_x = \frac{1}{d}, \qquad \pi_x = \frac{1}{2}, \qquad \mu_{1x} = \frac{1+\theta_x}{2}, \qquad \mu_{0x} = \frac{1-\theta_x}{2},$$

with

$$\Psi\{\mathrm{Obs}(P_\theta^{\mathrm{dense}})\} = \frac{1}{2} + \frac{1}{2d}\sum_{x=1}^{d} |\theta_x|.$$

- ***(Moment-matched priors and mixtures.)*** *The two dense product priors $\Pi_{0,n,d}$ and $\Pi_{1,n,d}$, with induced Poissonized mixtures $\mathbb{M}_{0,n,d}$ and $\mathbb{M}_{1,n,d}$, may be chosen so that their prior-mean separation satisfies*

$$\Delta_{n,d}^{\mathrm{dense}} = a_{n,d} E_{K_d^{\mathrm{lb}}}, \qquad 0 < \Delta_{n,d}^{\mathrm{dense}} \leq 1,$$

and

$$\mathrm{TV}(\mathbb{M}_{0,n,d}, \mathbb{M}_{1,n,d}) \leq d\{\mathrm{tail}_{n,d}^{\mathrm{dense}}\}^{1/2} \leq d\left\{\frac{(e/64)^{K_d^{\mathrm{lb}}+1}}{1-e/64}\right\}^{1/2} \leq \frac{1}{16}.$$

Under each of the two product priors, the target is concentrated around its prior mean:

$$\int \left(\Psi\{\mathrm{Obs}(P_\theta^{\mathrm{dense}})\} - \mathbb{E}_{\Pi_{j,n,d}}\Psi\{\mathrm{Obs}(P_\theta^{\mathrm{dense}})\}\right)^2 d\Pi_{j,n,d}(\theta) \le \frac{a_{n,d}^2}{4d}, \qquad j \in \{0,1\},$$

and

$$\Pi_{j,n,d}\left(\left|\Psi\{\mathrm{Obs}(P_\theta^{\mathrm{dense}})\} - \mathbb{E}_{\Pi_{j,n,d}}\Psi\{\mathrm{Obs}(P_\theta^{\mathrm{dense}})\}\right| > \frac{\Delta_{n,d}^{\mathrm{dense}}}{4}\right) \le \frac{1}{8}, \qquad j \in \{0,1\}.$$

- ***(One-cell likelihood identity.)*** *For all contrasts $\theta, \theta' \in \mathbb{R}$,*

$$E_0[L_\theta L_{\theta'}] = \exp(\lambda_{n,d}\theta\theta').$$

Consequently,

$$c_\epsilon(\Delta_{n,d}^{\mathrm{dense}})^2 \le \mathfrak{R}_{2n,d,\epsilon}^{\mathrm{Pois}}, \qquad c_\epsilon \frac{d}{nK_d^{\mathrm{lb}}} \le \mathfrak{R}_{2n,d,\epsilon}^{\mathrm{Pois}},$$

and the fixed-sample and Poissonized risks satisfy

$$\mathfrak{R}_{n,d,\epsilon} \ge \mathfrak{R}_{2n,d,\epsilon}^{\mathrm{Pois}} - \Pr\{\mathrm{Pois}(2n) < n\}.$$

In particular,

$$c_\epsilon \frac{d}{nL_d} \le \mathfrak{R}_{n,d,\epsilon}.$$

Proof of Lemma 10. Let

$$L_d = \log(ed), \qquad K_d^{\mathrm{lb}} = 2\lceil 4L_d \rceil, \qquad \lambda_{n,d} = \frac{2n}{d}, \qquad a_{n,d} = \sqrt{\frac{K_d^{\mathrm{lb}}}{64\lambda_{n,d}}}.$$

By Lemma 9, the standing conditions $2 \le d$ and $d^2 < n$ give

$$0 < a_{n,d} \le \frac{1}{2}.$$

The ceiling bounds also give

$$8L_d \le K_d^{\mathrm{lb}} < 8L_d + 2.$$

If $d \ge 2$, then $L_d = 1 + \log d \le d$. Hence

$$4L_d \le 4d, \qquad \lceil 4L_d \rceil \le 4d,$$

because $4d$ is an integer above $4L_d$. Therefore $K_d^{\mathrm{lb}} \le 8d$. If also $d^2 < n$, then $\lambda_{n,d} > 0$, and

$$a_{n,d}^2 = \frac{K_d^{\mathrm{lb}}}{64\lambda_{n,d}} = \frac{K_d^{\mathrm{lb}} d}{128n}.$$

[5]**Formalization scope.** This result uses the published conclusion [Cai and Low, 2011] (Lemma 1 and Section 3.1). The cited source proof is not formalized here: Lean verifies its use and all remaining steps. If that cited conclusion is false, the portions of this result that depend on it are not certified.

These relations, together with the amplitude range supplied above, prove the first group of claims.

For $\theta \in [-a_{n,d}, a_{n,d}]^d$, the amplitude bound places every coordinate in $[-1/2, 1/2]$. The dense law of Definition 24 has, in every cell,

$$p_x = \frac{1}{d}, \qquad \pi_x = \frac{1}{2}, \qquad \mu_{1x} = \frac{1+\theta_x}{2}, \qquad \mu_{0x} = \frac{1-\theta_x}{2},$$

equivalently

$$q_{11,x} = q_{00,x} = \frac{1+\theta_x}{4d}, \qquad q_{10,x} = q_{01,x} = \frac{1-\theta_x}{4d}.$$

Since $0 < \epsilon < 1/2$, the propensity $1/2$ belongs to $[\epsilon, 1-\epsilon]$, and the displayed means belong to $[0,1]$. Thus the observed margin lies in the class of Definition 3. By Definition 4,

$$\Psi\{\mathrm{Obs}(P_\theta^{\mathrm{dense}})\} = \frac{1}{d}\sum_{x=1}^{d} \max\left\{\frac{1+\theta_x}{2}, \frac{1-\theta_x}{2}\right\} = \frac{1}{2} + \frac{1}{2d}\sum_{x=1}^{d} |\theta_x|.$$

By [Cai and Low, 2011, Lemma 1 and Section 3.1], for every positive even K there are symmetric probability measures $\nu_{0,K}$ and $\nu_{1,K}$ on $[-1,1]$, with matching moments through degree K, such that

$$\int |t|\, d\nu_{1,K}(t) - \int |t|\, d\nu_{0,K}(t) = 2E_K, \qquad \frac{c_{\mathrm{CL}}}{K} \le E_K \le \frac{C_{\mathrm{CL}}}{K}$$

for universal constants $0 < c_{\mathrm{CL}} < C_{\mathrm{CL}} < \infty$. Choose these priors at $K = K_d^{\mathrm{lb}}$, take their d-fold products, and map each coordinate by $t \mapsto a_{n,d}t$. The preceding display for Ψ gives

$$\mathbb{E}_{\Pi_{j,n,d}} \Psi\{\mathrm{Obs}(P_\theta^{\mathrm{dense}})\} = \frac{1}{2} + \frac{a_{n,d}}{2d}\sum_{x=1}^{d} \int |t|\, d\nu_{j,K_d^{\mathrm{lb}}}(t),$$

and hence

$$\Delta_{n,d}^{\mathrm{dense}} = a_{n,d} E_{K_d^{\mathrm{lb}}}.$$

The lower Cai–Low bound gives $E_{K_d^{\mathrm{lb}}} > 0$. The zero polynomial is an admissible degree-K_d^{lb} approximant to $|t|$, and

$$\sup_{t \in [-1,1]} \big||t| - 0\big| \le 1,$$

so $E_{K_d^{\mathrm{lb}}} \le 1$. Together with $0 < a_{n,d} \le 1/2$, this yields

$$0 < \Delta_{n,d}^{\mathrm{dense}} \le 1.$$

The same reciprocal approximation bound supplies a universal cutoff D_{scale} such that, for every $d \ge D_{\mathrm{scale}}$,

$$32 \le d\, E_{K_d^{\mathrm{lb}}}^2.$$

Indeed, $K_d^{\mathrm{lb}} \le 10 L_d$ for $d \ge 2$, and $L_d^2 = o(\sqrt{d})$. Thus, after increasing D_{scale},

$$(K_d^{\mathrm{lb}})^2 \le 100\sqrt{e}\,\sqrt{d}, \qquad 3200\sqrt{e} \le c_{\mathrm{CL}}^2 \sqrt{d},$$

and therefore

$$dE_{K_d^{\mathrm{lb}}}^2 \ge d\frac{c_{\mathrm{CL}}^2}{(K_d^{\mathrm{lb}})^2} \ge \frac{c_{\mathrm{CL}}^2\sqrt{d}}{100\sqrt{e}} \ge 32.$$

For the one-cell likelihood identity, let R_+ and R_- be independent $\mathrm{Pois}(\lambda_{n,d}/2)$ counts under the zero-contrast baseline and define

$$L_\theta(R_+, R_-) = (1+\theta)^{R_+}(1-\theta)^{R_-}.$$

The Poisson probability-generating identity

$$\int c^r \, d\,\mathrm{Pois}(\rho)(r) = \exp\{\rho(c-1)\}$$

gives, for all $\theta, \theta' \in \mathbb{R}$,

$$E_0[L_\theta L_{\theta'}] = \exp\left[\frac{\lambda_{n,d}}{2}\{(1+\theta)(1+\theta')-1\} + \frac{\lambda_{n,d}}{2}\{(1-\theta)(1-\theta')-1\}\right] = \exp(\lambda_{n,d}\theta\theta').$$

Let $\rho_{j,n,d}$ be the one-cell mixture density relative to the zero-contrast sign-count baseline $Q_{0,n,d} = \mathrm{Pois}(\lambda_{n,d}/2) \otimes \mathrm{Pois}(\lambda_{n,d}/2)$:

$$\rho_{j,n,d}(r_+, r_-) = \int L_{a_{n,d}t}(r_+, r_-)\, d\nu_{j,K_d^{\mathrm{lb}}}(t), \qquad j \in \{0,1\}.$$

Expanding the exponential kernel above and using moment agreement through K_d^{lb} gives

$$\int (\rho_{1,n,d} - \rho_{0,n,d})^2 \, dQ_{0,n,d} \le 4 \sum_{r=K_d^{\mathrm{lb}}+1}^{\infty} \frac{(\lambda_{n,d} a_{n,d}^2)^r}{r!} = 4 \ \mathrm{tail}_{n,d}^{\mathrm{dense}}.$$

Define the supported one-cell sign kernel on $\mathbb{R}$ by

$$\mathsf{K}_{n,d}(t,\cdot) = \begin{cases} \mathrm{Law}(\mathrm{Pois}\{\lambda_{n,d}(1+a_{n,d}t)/2\}, \mathrm{Pois}\{\lambda_{n,d}(1-a_{n,d}t)/2\}), & |t| \le 1, \\ Q_{0,n,d}, & |t| > 1, \end{cases}$$

with independent coordinates in the first line. Its d-cell product prior-predictive law is

$$\mathfrak{S}_{j,n,d} = \int \bigotimes_{x=1}^{d} \mathsf{K}_{n,d}(t_x, \cdot)\, d\nu_{j,K_d^{\mathrm{lb}}}^{\otimes d}(t_1, \ldots, t_d), \qquad j \in \{0,1\}.$$

The Cauchy–Schwarz total-variation bound for one cell and the telescoping inequality for products yield

$$\mathrm{TV}(\mathfrak{S}_{0,n,d}, \mathfrak{S}_{1,n,d}) \le d\{\mathrm{tail}_{n,d}^{\mathrm{dense}}\}^{1/2}.$$

Let Sign map a Poisson observed sample to the array whose x-th entry is the pair of counts with $A = Y$ and $A \ne Y$ in cell x. Under the zero-contrast dense law, let

$$\mathsf{R}_{n,d}(r, \cdot) = \Pr_0\{\text{Poisson observed sample} \in \cdot \mid \mathrm{Sign} = r\}$$

be a regular conditional law. For every supported product prior above, this reconstruction kernel satisfies

$$\mathbb{M}_{j,n,d}(\mathcal{B}) = \int \mathsf{R}_{n,d}(r, \mathcal{B})\, d\mathfrak{S}_{j,n,d}(r), \qquad j \in \{0,1\},$$

for every measurable observed-sample event $\mathcal{B}$. Since total variation weakly decreases under a common Markov kernel,

$$\mathrm{TV}(\mathbb{M}_{0,n,d}, \mathbb{M}_{1,n,d}) \le \mathrm{TV}(\mathfrak{S}_{0,n,d}, \mathfrak{S}_{1,n,d}) \le d\{\mathrm{tail}_{n,d}^{\mathrm{dense}}\}^{1/2}.$$

Since

$$\lambda_{n,d} a_{n,d}^2 = \frac{K_d^{\mathrm{lb}}}{64},$$

Stirling's lower bound implies, for every $r \ge K_d^{\mathrm{lb}} + 1$,

$$\frac{(\lambda_{n,d} a_{n,d}^2)^r}{r!} = \frac{(K_d^{\mathrm{lb}}/64)^r}{r!} \le \left(\frac{e}{64}\right)^r.$$

Therefore

$$\mathrm{tail}_{n,d}^{\mathrm{dense}} \le \sum_{r=K_d^{\mathrm{lb}}+1}^{\infty} \left(\frac{e}{64}\right)^r = \frac{(e/64)^{K_d^{\mathrm{lb}}+1}}{1 - e/64}.$$

Put $\zeta_{\mathrm{geom}} = e/64$. Since $\zeta_{\mathrm{geom}} \le e^{-2}$, $K_d^{\mathrm{lb}} \ge 8L_d$, and $L_d = \log(ed)$,

$$\zeta_{\mathrm{geom}}^{K_d^{\mathrm{lb}}+1} \le e^{-2(K_d^{\mathrm{lb}}+1)} \le e^{-16L_d} = (ed)^{-16} \le \frac{1}{512d^2} \qquad (d \ge 2).$$

Also $1 - \zeta_{\mathrm{geom}} \ge 1/2$, and so

$$\frac{\zeta_{\mathrm{geom}}^{K_d^{\mathrm{lb}}+1}}{1 - \zeta_{\mathrm{geom}}} \le \frac{1}{256d^2}, \qquad d\left\{\frac{\zeta_{\mathrm{geom}}^{K_d^{\mathrm{lb}}+1}}{1 - \zeta_{\mathrm{geom}}}\right\}^{1/2} \le \frac{1}{16}.$$

Combining the last displays proves the total-variation chain in the statement.

Under either product prior, write $\theta_x = a_{n,d}\Xi_x$, where $\Xi_1, \ldots, \Xi_d$ are independent with common law $\nu_{j,K_d^{\mathrm{lb}}}$. The support condition makes the restriction to $[-1,1]^d$ immaterial, and

$$\Psi\{\mathrm{Obs}(P_\theta^{\mathrm{dense}})\} = \frac{1}{2} + \frac{a_{n,d}}{2d} \sum_{x=1}^{d} |\Xi_x|.$$

Because $|\Xi_x| \in [0,1]$, its variance is at most $1/4$. Independence gives

$$\int \Big(\Psi\{\mathrm{Obs}(P_\theta^{\mathrm{dense}})\} - \mathbb{E}_{\Pi_{j,n,d}}\Psi\{\mathrm{Obs}(P_\theta^{\mathrm{dense}})\}\Big)^2 \, d\Pi_{j,n,d}(\theta) \le \frac{a_{n,d}^2}{4d}.$$

Chebyshev's inequality, $\Delta_{n,d}^{\mathrm{dense}} = a_{n,d}E_{K_d^{\mathrm{lb}}}$, and $32 \le dE_{K_d^{\mathrm{lb}}}^2$ then yield

$$\Pi_{j,n,d}\left(\Big|\Psi\{\mathrm{Obs}(P_\theta^{\mathrm{dense}})\} - \mathbb{E}_{\Pi_{j,n,d}}\Psi\{\mathrm{Obs}(P_\theta^{\mathrm{dense}})\}\Big| > \frac{\Delta_{n,d}^{\mathrm{dense}}}{4}\right) \le \frac{a_{n,d}^2/(4d)}{(a_{n,d}E_{K_d^{\mathrm{lb}}}/4)^2} = \frac{4}{dE_{K_d^{\mathrm{lb}}}^2} \le \frac{1}{8}.$$

For any estimator in the Poissonized experiment, set

$$\bar{\psi}_j = \mathbb{E}_{\Pi_{j,n,d}}\Psi\{\mathrm{Obs}(P_\theta^{\mathrm{dense}})\}, \qquad r_\star = \max_{j\in\{0,1\}} \int E_\theta\left[\left\{\widehat{V} - \Psi\{\mathrm{Obs}(P_\theta^{\mathrm{dense}})\}\right\}^2\right] d\Pi_{j,n,d}(\theta).$$

Use the test

$$\varphi = \mathbf{1}\left\{\widehat{V} \geq \frac{\bar{\psi}_0 + \bar{\psi}_1}{2}\right\}, \qquad \alpha_0 = \mathbb{M}_{0,n,d}(\varphi = 1), \qquad \alpha_1 = \mathbb{M}_{1,n,d}(\varphi = 0).$$

On the concentration events above, estimation error smaller than $\Delta_{n,d}^{\mathrm{dense}}/4$ forces the test to select the correct prior index. Markov's inequality gives

$$\alpha_0 + \alpha_1 \leq \frac{1}{4} + \frac{32 r_\star}{(\Delta_{n,d}^{\mathrm{dense}})^2}.$$

Every test also satisfies

$$\alpha_0 + \alpha_1 \geq 1 - \mathrm{TV}(\mathbb{M}_{0,n,d}, \mathbb{M}_{1,n,d}) \geq \frac{15}{16}.$$

Consequently

$$\frac{11}{512}(\Delta_{n,d}^{\mathrm{dense}})^2 \leq \mathfrak{R}_{2n,d,\epsilon}^{\mathrm{Pois}}.$$

The restriction to dense observed laws is legitimate because those laws lie in $\mathcal{D}_{d,\epsilon}^{\mathrm{obs}}$. In addition, $\Psi(\mathbb{P}) \in [0,1]$ for every $\mathbb{P} \in \mathcal{D}_{d,\epsilon}^{\mathrm{obs}}$, since $p_x \geq 0$, $\sum_x p_x = 1$, and $0 \leq \max\{\mu_{0x}, \mu_{1x}\} \leq 1$; therefore projection of estimators onto $[0,1]$ weakly decreases squared loss.

The Cai–Low lower bound on E_K, together with the amplitude identity, gives a universal constant

$$c_0 = \frac{11 c_{\mathrm{CL}}^2}{512 \cdot 128} > 0$$

such that

$$c_0 \frac{d}{n K_d^{\mathrm{lb}}} \leq \frac{11}{512}(\Delta_{n,d}^{\mathrm{dense}})^2.$$

Indeed,

$$(\Delta_{n,d}^{\mathrm{dense}})^2 = \frac{K_d^{\mathrm{lb}} d}{128 n} E_{K_d^{\mathrm{lb}}}^2 \geq \frac{c_{\mathrm{CL}}^2}{128} \frac{d}{n K_d^{\mathrm{lb}}}.$$

Hence

$$c_0 \frac{d}{n K_d^{\mathrm{lb}}} \leq \mathfrak{R}_{2n,d,\epsilon}^{\mathrm{Pois}}.$$

It remains to compare the fixed-sample and Poissonized experiments. Given a fixed-n estimator, project it onto $[0,1]$, apply it to the first n observations of a mean-$2n$ Poisson sample on the event $\{\mathrm{Pois}(2n) \geq n\}$, and output 0 on the complementary event. On the first event, the first n observations have the n-fold i.i.d. law from Assumption 1; on the second event, the projected squared loss is at most 1. Taking infima gives

$$\mathfrak{R}_{n,d,\epsilon} \geq \mathfrak{R}_{2n,d,\epsilon}^{\mathrm{Pois}} - \Pr\{\mathrm{Pois}(2n) < n\}.$$

For the lower-tail absorption, fix any $\eta > 0$. Since $L_d = o(d)$, there is $D_{\log}$ such that

$$L_d \leq \eta d \qquad (d \geq D_{\log}).$$

Increasing the cutoff to ensure $d \geq 20$, the dense condition $d^2 < n$ gives $n \geq 400$. The mean-$2n$ Poisson Chernoff bound gives

$$\Pr\{\mathrm{Pois}(2n) < n\} \leq \exp\{-n(1 - \log 2)\}.$$

Since $1 - \log 2 > 3/10$ and $\log n \le 2\sqrt{n}$ for $n \ge 400$,

$$2 \log n \le n(1 - \log 2),$$

and therefore

$$\exp\{-n(1 - \log 2)\} \le n^{-2}.$$

For $d \ge D_{\log}$ and $n \ge 1$,

$$n^{-2} \le \frac{\eta d}{n L_d},$$

because $L_d \le \eta d n$. Thus, with $\eta = c_0/20$, there is a cutoff D_{tail} such that

$$\Pr\{\mathrm{Pois}(2n) < n\} \le \frac{c_0}{20} \frac{d}{n L_d}$$

throughout the dense regime above that cutoff.

Choose

$$D_0 = \max\{2, D_{\mathrm{scale}}, D_{\mathrm{tail}}\}, \qquad c_\epsilon = \min\left\{\frac{11}{512}, c_0, \frac{c_0}{20}\right\}.$$

Then $D_0 \ge 2$ and $c_\epsilon > 0$. The construction, observed-submodel identities, moment-matched priors, concentration bounds, total-variation chain, and one-cell likelihood identity have already been established for every $d \ge D_0$, $d^2 < n$, and $0 < \epsilon < 1/2$. Since $c_\epsilon \le 11/512$,

$$c_\epsilon (\Delta_{n,d}^{\mathrm{dense}})^2 \le \mathfrak{R}^{\mathrm{Pois}}_{2n,d,\epsilon}.$$

Since $c_\epsilon \le c_0$,

$$c_\epsilon \frac{d}{n K_d^{\mathrm{lb}}} \le \mathfrak{R}^{\mathrm{Pois}}_{2n,d,\epsilon}.$$

Finally, $K_d^{\mathrm{lb}} \le 10 L_d$, so

$$\frac{c_0}{10} \frac{d}{n L_d} \le c_0 \frac{d}{n K_d^{\mathrm{lb}}} \le \mathfrak{R}^{\mathrm{Pois}}_{2n,d,\epsilon}.$$

Using the Poisson-tail comparison and the fixed-sample transfer,

$$\frac{c_0}{20} \frac{d}{n L_d} \le \mathfrak{R}^{\mathrm{Pois}}_{2n,d,\epsilon} - \Pr\{\mathrm{Pois}(2n) < n\} \le \mathfrak{R}_{n,d,\epsilon}.$$

Because $c_\epsilon \le c_0/20$, this proves

$$c_\epsilon \frac{d}{n L_d} \le \mathfrak{R}_{n,d,\epsilon}.$$

All displayed conclusions follow. $\square$

The lemma supplies the dense-regime lower bound used by Theorem 2. Its first clauses place the submodel inside the fixed-overlap class of Definition 3 and identify the target as an absolute-value average. The prior clauses then combine the $|t|$ approximation gap with a total-variation bound for the induced mixtures, matching the reduction for estimating nonsmooth L_1-type functionals [Cai and Low, 2011, Jiao et al., 2018]. The final comparison transfers the Poissonized lower bound to the fixed-sample risk up to the displayed lower-tail term, yielding the $d/\{n \log(ed)\}$ rate contribution in the superquadratic-sample range.

C Verification note

Scope of the formal layer

The formal layer of the paper consists of the fifty-one displayed assumptions, definitions, algorithm, propositions, lemmas, remarks, and theorems that appear in the main text and appendices. Forty of them are attached to a declaration in a machine-checked Lean development: four assumptions, sixteen definitions and one algorithm, ten lemmas, four propositions, one remark, and four theorems. Every displayed theorem, proposition, and lemma belongs to this group, so each such statement is the reader-facing rendering of a Lean declaration whose proof is checked by the Lean kernel, subject to the published inputs recorded below. The proofs printed in this paper are prose renderings of those Lean proofs; they are not themselves the checked objects.

The remaining eleven displayed definitions are presentation-level: they name objects used in the exposition and carry no separate Lean declaration. They are the interval projection, the fixed-law squared risk, the fixed-sample two-sample L_1 minimax risk, the equal-propensity normalized L_1 embedding, the causal minimax risk, the tensor Jackson convolution operator, the good-pilot event, the pilot-failure contributions, the total-variation distance, the dense likelihood tail, and the one-cell dense likelihood ratio. Each abbreviates a quantity that the checked statements express directly in their own terms; none carries a claim beyond the checked statements that use it.

Artifact, toolchain, and commit

The Lean development for this paper is the directory `CausalSmith/Stat/STAT_DiscreteOptimalValueMinimaxMa` of the CausalSmith repository, at commit `881a4d27aaf42f02b1f71128db9f0c6d2184c790`. It is built with Lean 4 toolchain `leanprover/lean4:v4.33.0` and Mathlib at revision `db584cd6d46c92f209a44c0f1c829460d327499d`, through the build command `lake -d CausalSmith build`. The declarations backing the displayed statements are named in the interactive companion to this paper. A source scan of the development finds no `sorry` and no added axiom, and the axiom audit of its main theorems reports only the standard Lean axioms `propext`, `Classical.choice`, and `Quot.sound`.

Objects anchoring each layer

The model and identification layer is anchored by Assumptions 1 to 4, Definitions 2 to 6 and 9, and Propositions 1 and 2. Together these statements specify the fixed-overlap observed experiment, the full-data causal completion class, and the equality between the observed optimal-regression value and the causal optimal-treatment value under the stated conditions.

The estimation layer is anchored by Definitions 10, 12 and 13, Algorithm 1, and Theorem 1. These statements define the empirical-ratio benchmark, the Jackson smoothing polynomial, the all-data Rao–Blackwellized estimator, the fixed-law squared risk, and the uniform upper bound for the estimator over the fixed-overlap observed-law class.

The lower-bound and comparison layer is anchored by Definitions 14 to 16 and 24, Proposition 3, Lemmas 9 and 10, and Theorem 2.

The combined minimax conclusions are anchored by Theorems 3 and 4, Proposition 4, Definition 18, and Remark 1. These statements give the matched fixed-sample rate, the sequence-level consistency and parametric-rate characterizations, the comparison with the predecessor bracket on the same formal class, and the stated route for the sharp-overlap-constant question.

Published inputs the certification is conditional on

Two analytic conclusions enter as published inputs rather than as checked steps: the polynomial-approximation and moment-matching construction for $|t|$ of Cai and Low [2011, Lemma 1 and Section 3.1], and the Poissonized two-sample L_1 minimax lower bound of Jiao et al. [2018, Theorem 3, equation (24)]. The theorem-local footnotes identify the precise published conclusion each affected statement uses. The Jackson approximation material of DeVore and Lorentz [1993] and the testing and minimax comparison tools of Le Cam [1986], Tsybakov [2009] supply the classical background for the constructions, whose uses in this paper are checked. Using those two published inputs, the Lean development certifies the lower-bound and matched-frontier results through the dependencies identified in the theorem-local footnotes and checks the remaining steps internally.

D Proofs of the main results

Proof of Proposition 1. Write $\mathcal{J} = \{0,1\} \times \{0,1\}$. For a four-vector u, set

$$s_a(u) = u_{a0} + u_{a1}, \qquad s(u) = s_0(u) + s_1(u), \qquad g_{a,\epsilon}(u) = \begin{cases} 0, & s(u) = 0, \\ \dfrac{s(u)u_{a1}}{\max\{s_a(u), \epsilon s(u)\}}, & s(u) > 0, \end{cases}$$

so that $\bar{f}_\epsilon(u) = \max\{g_{0,\epsilon}(u), g_{1,\epsilon}(u)\}$ on $\mathbb{R}^4_+$, as in Definition 8.

1. Fix $x \in [d]$. The coordinates of q_x are observed atom masses, hence $q_{\jmath,x} \geq 0$ for every $\jmath \in \mathcal{J}$. Moreover
$$s(q_x) = p_x, \qquad s_1(q_x) = q_{10,x} + q_{11,x} = p_x \pi_x.$$
If $p_x > 0$, Definition 3 gives
$$\epsilon \leq \frac{s_1(q_x)}{s(q_x)} \leq 1 - \epsilon,$$
and therefore
$$\epsilon s(q_x) \leq s_1(q_x) \leq (1-\epsilon)s(q_x).$$
If $p_x = 0$, nonnegativity and $s(q_x) = 0$ force $q_x = 0$, so the same two inequalities hold. Thus $q_x \in \mathcal{C}_\epsilon$ by Definition 7.

2. For each arm $a \in \{0,1\}$, define the totalized cell mean
$$\mu^{\text{tot}}_{ax} = \begin{cases} q_{a1,x}/s_a(q_x), & s_a(q_x) > 0, \\ 0, & s_a(q_x) = 0. \end{cases}$$
On an occupied cell, the cone inequalities from the previous step imply
$$s_1(q_x) \geq \epsilon s(q_x) > 0, \qquad s_0(q_x) = s(q_x) - s_1(q_x) \geq \epsilon s(q_x) > 0,$$
and μ^{tot}_{ax} agrees with the outcome mean μ_{ax} in Definition 3. Hence, for both arms,
$$\max\{s_a(q_x), \epsilon s(q_x)\} = s_a(q_x), \qquad q_{a1,x} = s_a(q_x)\mu^{\text{tot}}_{ax},$$

and therefore

$$g_{a,\epsilon}(q_x) = \frac{s(q_x)q_{a1,x}}{s_a(q_x)} = p_x\mu_{ax}^{\mathrm{tot}}.$$

On a null cell all observed atom masses vanish, and both sides of this last identity are zero. Taking the maximum over a gives

$$\bar{f}_\epsilon(q_x) = p_x \max\{\mu_{0x}, \mu_{1x}\},$$

with the displayed value unchanged by the permitted null-cell choice of μ_{ax}. Summing over $x \in [d]$ and using Definition 4 yields

$$\Psi(\mathbb{P}) = \sum_{x=1}^{d} \bar{f}_\epsilon(q_x).$$

3. Let $u \in \mathbb{R}_+^4$. For every arm a, $0 \le u_{a1} \le s_a(u) \le \max\{s_a(u), \epsilon s(u)\}$. If $s(u) = 0$, then $g_{a,\epsilon}(u) = 0$. If $s(u) > 0$, the denominator is positive and

$$0 \le g_{a,\epsilon}(u) = \frac{s(u)u_{a1}}{\max\{s_a(u), \epsilon s(u)\}} \le s(u).$$

Taking the maximum over the two arms gives

$$0 \le \bar{f}_\epsilon(u) \le s(u).$$

4. It remains to prove the Lipschitz estimate for the global extension. For nonnegative $\phi, \zeta, \rho, \upsilon$, put

$$\mathrm{scal}_\epsilon(\phi,\zeta,\rho,\upsilon) = \begin{cases} 0, & \phi+\zeta+\rho+\upsilon = 0, \\ \dfrac{(\phi+\zeta+\rho+\upsilon)\zeta}{\max\{\phi+\zeta, \epsilon(\phi+\zeta+\rho+\upsilon)\}}, & \phi+\zeta+\rho+\upsilon > 0. \end{cases}$$

For two quadruples, write

$$\mathrm{dist}_{\mathrm{cell}} = |\phi-\phi'|+|\zeta-\zeta'|+|\rho-\rho'|+|\upsilon-\upsilon'|, \qquad \mathrm{tot} = \phi+\zeta+\rho+\upsilon, \quad \mathrm{tot}' = \phi'+\zeta'+\rho'+\upsilon'.$$

When both totals are positive, split according to whether $\phi+\zeta \le \epsilon\,\mathrm{tot}$ and $\phi'+\zeta' \le \epsilon\,\mathrm{tot}'$. In the low–low case,

$$|\,\mathrm{scal}_\epsilon - \mathrm{scal}'_\epsilon\,| = \left|\frac{\zeta-\zeta'}{\epsilon}\right| \le \epsilon^{-1}\,\mathrm{dist}_{\mathrm{cell}}\,.$$

In the low–high case,

$$\frac{\zeta}{\epsilon} - \mathrm{tot}'\,\frac{\zeta'}{\phi'+\zeta'} = \frac{\zeta-\zeta'}{\epsilon} + \frac{\zeta'}{\phi'+\zeta'}\,\frac{(\phi'+\zeta')-\epsilon\,\mathrm{tot}'}{\epsilon}.$$

The second summand is nonnegative, and $0 \le \zeta'/(\phi'+\zeta') \le 1$. Using $\phi+\zeta \le \epsilon\,\mathrm{tot}$ gives

$$\frac{\zeta-\zeta'}{\epsilon} + \frac{(\phi'+\zeta')-\epsilon\,\mathrm{tot}'}{\epsilon} \le \epsilon^{-1}|\phi-\phi'| + |\,\mathrm{tot}-\mathrm{tot}'\,| \le (1+\epsilon^{-1})\,\mathrm{dist}_{\mathrm{cell}},$$

and the lower bound follows from

$$\frac{\zeta - \zeta'}{\epsilon} \geq -\epsilon^{-1} \operatorname{dist}_{\text{cell}} .$$

The high–low case is the same calculation with the two quadruples interchanged. In the high–high case,

$$\operatorname{tot} \frac{\zeta}{\phi + \zeta} - \operatorname{tot}' \frac{\zeta'}{\phi' + \zeta'} = (\operatorname{tot} - \operatorname{tot}') \frac{\zeta'}{\phi' + \zeta'} + \operatorname{tot} \left(\frac{\zeta}{\phi + \zeta} - \frac{\zeta'}{\phi' + \zeta'} \right),$$

and

$$\left| \frac{\zeta}{\phi + \zeta} - \frac{\zeta'}{\phi' + \zeta'} \right| \leq \frac{|\zeta - \zeta'| + |\phi - \phi'|}{\phi + \zeta},$$

because

$$\frac{\zeta}{\phi + \zeta} - \frac{\zeta'}{\phi' + \zeta'} = \frac{\phi'(\zeta - \zeta') + \zeta'(\phi' - \phi)}{(\phi + \zeta)(\phi' + \zeta')}.$$

Together with $0 \leq \zeta'/(\phi' + \zeta') \leq 1$ and $\operatorname{tot}/(\phi + \zeta) \leq \epsilon^{-1}$, this gives

$$|\operatorname{scal}_\epsilon - \operatorname{scal}'_\epsilon| \leq (1 + \epsilon^{-1}) \operatorname{dist}_{\text{cell}} .$$

If one total is zero, nonnegativity forces that quadruple to vanish, and the bound follows from $0 \leq \operatorname{scal}_\epsilon \leq \operatorname{tot}$ and $1 \leq 1 + \epsilon^{-1}$. Applying this estimate to each arm, after merely permuting the four coordinates, yields

$$|g_{a,\epsilon}(u) - g_{a,\epsilon}(v)| \leq (1 + \epsilon^{-1}) \sum_{\jmath \in \mathcal{J}} |u_\jmath - v_\jmath|.$$

Finally,

$$|\bar{f}_\epsilon(u) - \bar{f}_\epsilon(v)| \leq \max_{a \in \{0,1\}} |g_{a,\epsilon}(u) - g_{a,\epsilon}(v)|,$$

which proves the asserted global Lipschitz property.

5. Define

$$\operatorname{bm}(\eta, \iota) = \begin{cases} \eta, & \iota = 1, \\ 1 - \eta, & \iota = 0, \end{cases} \qquad \eta \in [0,1],\ \iota \in \{0,1\}.$$

Construct a full-data mass function by

$$P^{\mathrm{B}}(x, a, y, y_0, y_1) = \mathbf{1}\{y = y_a\}\, q_{ay,x}\, \operatorname{bm}(\mu_{1-a,x}, y_{1-a}),$$

where $y_a = y_1$ for $a = 1$ and $y_a = y_0$ for $a = 0$. Since $\mu_{0x}, \mu_{1x} \in [0,1]$, every displayed mass is nonnegative. Also

$$\sum_{x,a,y,y_0,y_1} P^{\mathrm{B}}(x, a, y, y_0, y_1) = \sum_{x,a,y} q_{ay,x} \sum_{y_{1-a} \in \{0,1\}} \operatorname{bm}(\mu_{1-a,x}, y_{1-a}) = \sum_{x,a,y} q_{ay,x} = 1.$$

Thus these masses define a full-data law.

The construction gives consistency because atoms with $y \neq y_a$ have mass zero. Its observed margin is $\mathbb{P}$, since

$$\sum_{y_0, y_1 \in \{0,1\}} P^{\mathrm{B}}(x, a, y, y_0, y_1) = q_{ay,x} \sum_{y_{1-a} \in \{0,1\}} \operatorname{bm}(\mu_{1-a,x}, y_{1-a}) = q_{ay,x}.$$

Using the observed-law factorization supplied by Definition 3,

$$q_{ay,x} = p_x \ \mathrm{bm}(\pi_x, a) \ \mathrm{bm}(\mu_{ax}, y),$$

the potential-outcome atom obtained after summing over the observed outcome factors as

$$\sum_{y\in\{0,1\}} P^{\mathrm{B}}(x, a, y, y_0, y_1) = p_x \ \mathrm{bm}(\pi_x, a) \ \mathrm{bm}(\mu_{0x}, y_0) \ \mathrm{bm}(\mu_{1x}, y_1).$$

Consequently, conditionally on $X = x$, treatment is Bernoulli with parameter π_x, independent of the joint potential-outcome vector, and $Y(0), Y(1)$ are independent Bernoulli variables with parameters μ_{0x} and μ_{1x}. The observed margin is $\mathbb{P}$, so its overlap properties are exactly those of the given observed law; together with consistency and conditional exchangeability, Definition 6 places P^{B} in $\mathcal{P}_{d,\epsilon}$.

6. Conversely, if $P \in \mathcal{P}_{d,\epsilon}$, then Definition 6 includes the assertion that its observed margin satisfies the observed fixed-overlap model restrictions. Hence

$$\mathrm{Obs}(P) \in \mathcal{D}^{\mathrm{obs}}_{d,\epsilon}.$$

The preceding conclusions use the nonnegative-domain extension $u \mapsto \bar{f}_\epsilon(u)$ from Definition 8, so they are exactly the cell-overlap, value, global-bound, Lipschitz, completion, and observed-margin assertions stated above.

□

Proof of Proposition 2. Let

$$\mathbb{P} := \mathrm{Obs}(P).$$

For $x \in [d]$ and $a \in \{0,1\}$, write

$$p_x := \mathbb{P}(X = x), \qquad \omega_{a,x} := \mathbb{P}(X = x, A = a),$$

and define the totalized observed arm regression by

$$\xi_{a,x} := \begin{cases} \dfrac{\mathbb{P}(X = x, A = a, Y = 1)}{\omega_{a,x}}, & \omega_{a,x} > 0, \\ 0, & \omega_{a,x} = 0. \end{cases}$$

Also write

$$\rho_x := P(X = x).$$

For $\iota \in \{0,1\}$ and $\tau, \upsilon \in \{0,1\}$, put

$$\chi_{\iota,x}(\tau, \upsilon) := P\{X = x, A = \tau, Y(\iota) = \upsilon\},$$

and define the totalized potential-outcome regression by

$$\eta_{\iota,x} := \begin{cases} \dfrac{\sum_{\tau\in\{0,1\}} \chi_{\iota,x}(\tau, 1)}{\rho_x}, & \rho_x > 0, \\ 0, & \rho_x = 0. \end{cases}$$

1. By Definition 2, for every x, a, y,

$$\mathbb{P}(X = x, A = a, Y = y) = P\{X = x, A = a, Y = y\}.$$

The support component of Definition 6 makes $Y(0)$ and $Y(1)$ binary, so the event $\{X = x, A = a, Y = y\}$ is the disjoint union, over $y_0, y_1 \in \{0, 1\}$, of the events

$$\{X = x, A = a, Y = y, Y(0) = y_0, Y(1) = y_1\}.$$

Finite additivity therefore gives

$$\mathbb{P}(X = x, A = a, Y = y) = \sum_{y_0 \in \{0,1\}} \sum_{y_1 \in \{0,1\}} P\{X = x, A = a, Y = y, Y(0) = y_0, Y(1) = y_1\}.$$

Summing over a and y gives

$$p_x = \rho_x.$$

2. Fix $x \in [d]$ and $\iota \in \{0, 1\}$, and suppose

$$p_x > 0, \qquad \omega_{\iota,x} > 0.$$

The consistency component of Definition 6 gives, for each $\upsilon \in \{0, 1\}$,

$$\mathbb{P}(X = x, A = \iota, Y = \upsilon) = \chi_{\iota,x}(\iota, \upsilon),$$

and hence

$$\omega_{\iota,x} = \sum_{\upsilon \in \{0,1\}} \chi_{\iota,x}(\iota, \upsilon).$$

The exchangeability component of Definition 6, applied to $Y(\iota)$, $A = \iota$, and outcome value 1, gives

$$\chi_{\iota,x}(\iota, 1) \sum_{\tau \in \{0,1\}} \sum_{\upsilon \in \{0,1\}} \chi_{\iota,x}(\tau, \upsilon) = \left(\sum_{\tau \in \{0,1\}} \chi_{\iota,x}(\tau, 1) \right) \left(\sum_{\upsilon \in \{0,1\}} \chi_{\iota,x}(\iota, \upsilon) \right).$$

The double sum is $\rho_x = p_x$. Dividing by the positive product $p_x \omega_{\iota,x}$ yields

$$\eta_{\iota,x} = \frac{\sum_{\tau \in \{0,1\}} \chi_{\iota,x}(\tau, 1)}{\rho_x} = \frac{\chi_{\iota,x}(\iota, 1)}{\omega_{\iota,x}} = \xi_{\iota,x}.$$

3. The support component of Definition 6 gives the finite partition $\{X = x\}_{x \in [d]}$. Expanding the outer expectation in Definition 9 over this partition gives

$$V^\star(P) = \sum_{x : \rho_x > 0} \rho_x \max \left\{ \frac{\sum_{\tau \in \{0,1\}} \chi_{0,x}(\tau, 1)}{\rho_x}, \frac{\sum_{\tau \in \{0,1\}} \chi_{1,x}(\tau, 1)}{\rho_x} \right\}.$$

By the defining convention for $\eta_{\iota,x}$ on null cells, this is equivalently

$$V^\star(P) = \sum_{x \in [d]} \rho_x \max\{\eta_{0,x}, \eta_{1,x}\}.$$

Since $P \in \mathcal{P}_{d,\epsilon}$, Definition 6 places $\mathbb{P}$ in the observed-law class of Definition 3. Choose the representing parameters p_x^{rep}, π_x, and μ_{ax} from that representation. Summing its displayed atom factorization over a and y gives

$$p_x^{\mathrm{rep}} = \mathbb{P}(X = x) = p_x,$$

and summing over y at fixed a gives

$$\omega_{a,x} = \mathbb{P}(X = x, A = a) = p_x \pi_x^a (1 - \pi_x)^{1-a}.$$

If $p_x > 0$, fixed overlap in Definition 3 gives $\omega_{a,x} > 0$, and hence

$$\xi_{a,x} = \frac{\mathbb{P}(X = x, A = a, Y = 1)}{\omega_{a,x}} = \frac{p_x \pi_x^a (1 - \pi_x)^{1-a} \mu_{ax}}{p_x \pi_x^a (1 - \pi_x)^{1-a}} = \mu_{ax}.$$

If $p_x = 0$, the cell contribution to the value is zero for any permitted choice of μ_{ax}. Therefore Definition 4 gives

$$\Psi(\mathbb{P}) = \sum_{x \in [d]} p_x \max\{\xi_{0,x}, \xi_{1,x}\}.$$

Using $p_x = \rho_x$, fix x. If $p_x = 0$, then

$$\rho_x \max\{\eta_{0,x}, \eta_{1,x}\} = p_x \max\{\xi_{0,x}, \xi_{1,x}\} = 0.$$

If $p_x > 0$, the overlap component of Definition 6 gives

$$\epsilon \le \frac{\omega_{1,x}}{p_x} \le 1 - \epsilon.$$

Since $\epsilon > 0$,

$$0 < \epsilon p_x \le \omega_{1,x}.$$

Also

$$p_x = \omega_{0,x} + \omega_{1,x}, \qquad \omega_{1,x} \le (1 - \epsilon) p_x,$$

so

$$\omega_{0,x} = p_x - \omega_{1,x} \ge \epsilon p_x > 0.$$

The identity from the previous step applies for $\iota = 0$ and $\iota = 1$, giving

$$\eta_{0,x} = \xi_{0,x}, \qquad \eta_{1,x} = \xi_{1,x}.$$

Thus

$$\rho_x \max\{\eta_{0,x}, \eta_{1,x}\} = p_x \max\{\xi_{0,x}, \xi_{1,x}\}.$$

Summing over $x \in [d]$ proves

$$V^\star(P) = \Psi\{\mathrm{Obs}(P)\}.$$

$\square$

Proof of Theorem 1. 1. Choose the universal tuning supplied by Lemma 8. Thus H_0 is the universal self-normalized Poisson pilot-radius constant,

$$\kappa = \frac{1}{100000}, \qquad D_0 = 3,$$

so $H_0 > 0$, $\kappa \in (0, 1)$, and $D_0 > 2$. Fix $0 < \epsilon < 1/2$, let $C_\epsilon^{\text{cell}} > 0$ be the cell constant supplied by Lemma 8, and put

$$\Xi_\epsilon := 1 + \epsilon^{-1} + \epsilon^{-2}.$$

Define

$$C_\epsilon := 20000000\{1 + C_\epsilon^{\text{cell}} + (C_\epsilon^{\text{cell}})^2\} + 2000\Xi_\epsilon + 10.$$

Then $C_\epsilon > 0$. For $d \geq 2$, write

$$L_d := \log(ed).$$

Since $L_d = 1 + \log d$ and $\log d \leq d - 1$,

$$0 < L_d \leq d, \qquad 1 \leq L_2 \leq 2.$$

Fix $n \geq 1$, $d \geq 2$, and $\mathbb{P} \in \mathcal{D}_{d,\epsilon}^{\text{obs}}$. Let $q_x = (q_{00,x}, q_{01,x}, q_{10,x}, q_{11,x})$ and let p_x be the covariate-cell mass. By Proposition 1,

$$\Psi(\mathbb{P}) = \sum_{x=1}^{d} \bar{f}_\epsilon(q_x), \qquad 0 \leq \bar{f}_\epsilon(q_x) \leq s(q_x).$$

The four-cell vectors partition the one-observation law, so

$$\sum_{x=1}^{d} s(q_x) = 1.$$

Consequently

$$0 \leq \Psi(\mathbb{P}) \leq 1.$$

2. Consider the bounded-alphabet branch $d < D_0$. Since $D_0 = 3$ and $d \geq 2$, this branch has $d = 2$. By Algorithm 1,

$$\widehat{V}_{n,2}^{\text{JF}} = \widehat{V}_{n,2}^{\text{ER}}.$$

Assume first that $n \geq 4$. For $x \in [2]$ and $a \in \{0, 1\}$, define

$$\zeta_{a,x} := \frac{\overline{N}_x}{n} \frac{\overline{S}_{ax}}{\overline{N}_{ax}}, \qquad \zeta_{a,x}^0 := p_x \mu_{ax},$$

with the $0/0 = 0$ convention from Definition 10, and put

$$\delta_{a,x} := \zeta_{a,x} - \zeta_{a,x}^0.$$

If $p_x = 0$, then $\zeta_{a,x}^0 = 0$ and $\zeta_{a,x} = 0$ almost surely. If $p_x > 0$, the fixed-overlap representation in Definition 3 gives

$$\eta_{a,x} := \mathbb{P}(A = a \mid X = x) \geq \epsilon,$$

and the conditional success mean in that arm and cell is μ_{ax}. Define

$$\delta_{a,x}^{\mathrm{res}} := \frac{\overline{N}_x}{n}\mathbf{1}\{\overline{N}_{ax} > 0\}\frac{1}{\overline{N}_{ax}}\sum_{i=1}^{n}\mathbf{1}\{X_i = x, A_i = a\}(Y_i - \mu_{ax}),$$

$$\delta_{a,x}^{\mathrm{cell}} := \left(\frac{\overline{N}_x}{n} - p_x\right)\mu_{ax}, \qquad \delta_{a,x}^{\mathrm{zero}} := \mu_{ax}\frac{\overline{N}_x}{n}\mathbf{1}\{\overline{N}_{ax} = 0\}.$$

The totalized-ratio identity is

$$\delta_{a,x} = \delta_{a,x}^{\mathrm{res}} + \delta_{a,x}^{\mathrm{cell}} - \delta_{a,x}^{\mathrm{zero}}.$$

Conditional on $\overline{N}_x = k$, the arm count is binomial with success probability $\eta_{a,x}$. For $\mathcal{B}_{k,\eta} \sim \mathrm{Bin}(k, \eta)$ with $\eta \geq \epsilon$,

$$E\left[\frac{\mathbf{1}\{\mathcal{B}_{k,\eta} > 0\}}{\mathcal{B}_{k,\eta}}\right] \leq 2E\left[\frac{1}{\mathcal{B}_{k,\eta} + 1}\right] = \frac{2\{1 - (1-\eta)^{k+1}\}}{(k+1)\eta} \leq \frac{2}{k\epsilon}$$

for $k > 0$, while the left-hand side is 0 for $k = 0$. Since $Y_i \in \{0, 1\}$ and $\mu_{ax} \in [0, 1]$, the centered residual has conditional second moment at most one; hence

$$E_{\mathbb{P}}\{(\delta_{a,x}^{\mathrm{res}})^2\} \leq E_{\mathbb{P}}\left[\frac{\overline{N}_x^2}{n^2}\frac{\mathbf{1}\{\overline{N}_{ax} > 0\}}{\overline{N}_{ax}}\right] \leq \frac{2p_x}{n\epsilon}.$$

Also

$$E_{\mathbb{P}}\{(\delta_{a,x}^{\mathrm{cell}})^2\} \leq E_{\mathbb{P}}\left(\frac{\overline{N}_x}{n} - p_x\right)^2 \leq \frac{1}{n}.$$

Let

$$\Lambda_{n,x} := p_x \exp\left[-\frac{n-2}{2}\epsilon p_x\right].$$

Expanding $\overline{N}_x^2\mathbf{1}\{\overline{N}_{ax} = 0\}$ into diagonal and ordered-pair contributions gives

$$E_{\mathbb{P}}\{(\delta_{a,x}^{\mathrm{zero}})^2\} \leq \frac{1}{n} + \Lambda_{n,x}^2.$$

The elementary inequality $t\exp(-ut) \leq u^{-1}$, applied with $u = (n-2)\epsilon/2$, and the condition $n \geq 4$ give

$$\Lambda_{n,x} \leq \frac{4}{n\epsilon}.$$

Using $p_x \leq 1$ and $n^{-2} \leq n^{-1}$, we obtain

$$E_{\mathbb{P}}\delta_{a,x}^2 \leq 4E_{\mathbb{P}}\{(\delta_{a,x}^{\mathrm{res}})^2\} + 2E_{\mathbb{P}}\{(\delta_{a,x}^{\mathrm{cell}})^2\} + 4E_{\mathbb{P}}\{(\delta_{a,x}^{\mathrm{zero}})^2\} \leq \frac{100\Xi_\epsilon}{n}.$$

Moreover,

$$\widehat{V}_{n,2}^{\mathrm{ER}} = \sum_{x=1}^{2}\max\{\zeta_{0,x}, \zeta_{1,x}\}, \qquad \Psi(\mathbb{P}) = \sum_{x=1}^{2}\max\{\zeta_{0,x}^0, \zeta_{1,x}^0\}.$$

The maximum map is one-Lipschitz in the sup norm, and Cauchy–Schwarz over the four pairs (a, x) yields

$$\left(\widehat{V}_{n,2}^{\mathrm{ER}} - \Psi(\mathbb{P})\right)^2 \le 4 \sum_{x=1}^{2} \sum_{a \in \{0,1\}} \delta_{a,x}^2.$$

Therefore

$$\mathrm{Risk}_{\mathbb{P}}(\widehat{V}_{n,2}^{\mathrm{ER}}, \Psi) \le \frac{1600 \Xi_\epsilon}{n}.$$

Since $n \ge 4$, $L_2 \ge 1$, and $L_2 \le 2$,

$$\min\left\{1, \frac{2}{nL_2}\right\} = \frac{2}{nL_2}, \qquad \frac{1600\Xi_\epsilon}{n} \le C_\epsilon \frac{2}{nL_2}.$$

If $n \le 3$, the empirical ratios lie in $[0, 1]$, their weights sum to one, and therefore

$$0 \le \widehat{V}_{n,2}^{\mathrm{ER}} \le 1.$$

Together with $0 \le \Psi(\mathbb{P}) \le 1$, this gives

$$\mathrm{Risk}_{\mathbb{P}}(\widehat{V}_{n,2}^{\mathrm{ER}}, \Psi) \le 1.$$

Also

$$\frac{1}{3} \le \frac{2}{nL_2}, \qquad \frac{1}{3} \le 1,$$

so

$$\frac{1}{3} \le \min\left\{1, \frac{2}{nL_2}\right\}.$$

Since $C_\epsilon \ge 3$, the desired bound follows throughout the bounded-alphabet branch.

3. Now suppose that $d \ge D_0$ and

$$n < \frac{d}{L_d}.$$

In this branch, Algorithm 1 outputs the constant estimator $1/2$. Since $\Psi(\mathbb{P}) \in [0, 1]$,

$$\mathrm{Risk}_{\mathbb{P}}\left(\frac{1}{2}, \Psi\right) \le 1.$$

The scale condition gives

$$1 < \frac{d}{nL_d},$$

and hence

$$\min\left\{1, \frac{d}{nL_d}\right\} = 1.$$

Since $C_\epsilon \ge 1$, the asserted inequality holds in this branch.

4. It remains to treat the Jackson branch

$$d \ge D_0, \qquad \frac{d}{L_d} \le n.$$

Set

$$m := \frac{n}{8}, \qquad \chi_{n,d} := \frac{d}{nL_d}.$$

Then $0 \le \chi_{n,d} \le 1$. In the uncapped finite-Poisson marked experiment, draw

$$M^{\mathrm{unc}} \sim \mathrm{Pois}(n/4),$$

then draw M^{unc} i.i.d. observations from $\mathbb{P}$, each with an independent fair pilot/evaluation mark. Let $T_x^{\mathrm{JF,unc}}$ be the clipped cell statistic computed from the resulting cell-x pilot and evaluation counts, and define

$$\widetilde{V}^{\mathrm{unc}} := \Pi_{[0,1]}\left(\sum_{x=1}^{d} T_x^{\mathrm{JF,unc}}\right).$$

By Lemma 8, the cell blocks are independent and, for every x,

$$\left|ET_x^{\mathrm{JF,unc}} - \bar{f}_\epsilon(q_x)\right| \le C_\epsilon^{\mathrm{cell}}\left\{\sqrt{\frac{p_x}{mL_d}} + \frac{1}{mL_d}\right\},$$

$$\mathrm{Var}(T_x^{\mathrm{JF,unc}}) \le C_\epsilon^{\mathrm{cell}} d^{1/16}\left\{\frac{p_x L_d}{m} + \frac{L_d^2}{m^2}\right\}.$$

Projection onto $[0,1]$ is nonexpansive for squared distance from $\Psi(\mathbb{P}) \in [0,1]$, so

$$E\left[\left\{\widetilde{V}^{\mathrm{unc}} - \Psi(\mathbb{P})\right\}^2\right] \le E\left[\left\{\sum_{x=1}^{d} T_x^{\mathrm{JF,unc}} - \sum_{x=1}^{d} \bar{f}_\epsilon(q_x)\right\}^2\right].$$

Using independence,

$$E\left[\left\{\sum_{x=1}^{d} T_x^{\mathrm{JF,unc}} - \sum_{x=1}^{d} \bar{f}_\epsilon(q_x)\right\}^2\right] \le 2\sum_{x=1}^{d}\mathrm{Var}(T_x^{\mathrm{JF,unc}}) + 2\left(\sum_{x=1}^{d} ET_x^{\mathrm{JF,unc}} - \sum_{x=1}^{d} \bar{f}_\epsilon(q_x)\right)^2.$$

Since $\sum_x p_x = 1$, Cauchy–Schwarz gives

$$\left|\sum_{x=1}^{d} ET_x^{\mathrm{JF,unc}} - \sum_{x=1}^{d} \bar{f}_\epsilon(q_x)\right|^2 \le 2(C_\epsilon^{\mathrm{cell}})^2\left\{\frac{d}{mL_d} + \frac{d^2}{m^2L_d^2}\right\}.$$

Summing the variance bounds gives

$$\sum_{x=1}^{d}\mathrm{Var}(T_x^{\mathrm{JF,unc}}) \le C_\epsilon^{\mathrm{cell}} d^{1/16}\left\{\frac{L_d}{m} + \frac{dL_d^2}{m^2}\right\}.$$

The elementary logarithmic absorption bounds

$$L_d^2 d^{-15/16} \le 225, \qquad L_d^4 d^{-15/16} \le 50625$$

hold for every integer $d \ge 2$: the one-variable bound $(1+\log t)^2 t^{-3/16} \le 225$ for $t \ge 1$ gives the first inequality after multiplication by $d^{-12/16} \le 1$, and the second after squaring and multiplication by $d^{-9/16} \le 1$. With $m = n/8$ and $\chi_{n,d} \le 1$, these imply

$$\frac{d}{mL_d} + \frac{d^2}{m^2L_d^2} \le 72\chi_{n,d},$$

and the scale inequality $d/L_d \le n$ gives

$$d^{1/16}\left\{\frac{L_d}{m}+\frac{dL_d^2}{m^2}\right\} \le 3300000\chi_{n,d}.$$

Consequently,

$$E\left[\left\{\widetilde{V}^{\mathrm{unc}}-\Psi(\mathbb{P})\right\}^2\right] \le 10000000\{1+C_\epsilon^{\mathrm{cell}}+(C_\epsilon^{\mathrm{cell}})^2\}\chi_{n,d}.$$

5. We pass from the uncapped comparison experiment to the fixed-sample estimator in Algorithm 1. Let $\mathbb{B}_n$ be the product law of n independent fair marks. For

$$o_{1:n} \in ([d]\times\{0,1\}^2)^n, \qquad \beta_{1:n}\in\{0,1\}^n, \qquad M\in\mathbb{N},$$

define

$$\Upsilon_{n,d}^{\mathrm{cap}}(o_{1:n},\beta_{1:n},M) := \begin{cases} \Pi_{[0,1]}\left\{\sum_{x=1}^d T_x^{\mathrm{JF}}(o_{1:M},\beta_{1:M})\right\}, & M\le n,\\ 0, & M>n.\end{cases}$$

Define the capped Poisson-prefix average

$$\overline{\Upsilon}_{n,d}^{\mathrm{cap}}(o_{1:n},\beta_{1:n}) := \int_{\mathbb{N}} \Upsilon_{n,d}^{\mathrm{cap}}(o_{1:n},\beta_{1:n},M)\, d\,\mathrm{Pois}_{n/4}(M).$$

In the present Jackson branch, the all-data estimator is the fair-mark average

$$\widehat{V}_{n,d}^{\mathrm{JF}}(o_{1:n}) = \int \overline{\Upsilon}_{n,d}^{\mathrm{cap}}(o_{1:n},\beta_{1:n})\, d\mathbb{B}_n(\beta_{1:n}).$$

Jensen's inequality for the mark average gives

$$\mathrm{Risk}_{\mathbb{P}}(\widehat{V}_{n,d}^{\mathrm{JF}},\Psi) \le E_{\mathbb{P}^{\otimes n}\otimes\mathbb{B}_n}\left[\left\{\overline{\Upsilon}_{n,d}^{\mathrm{cap}}(O_{1:n},B_{1:n})-\Psi(\mathbb{P})\right\}^2\right].$$

A second application of Jensen, now to the Poisson-prefix average, gives

$$E_{\mathbb{P}^{\otimes n}\otimes\mathbb{B}_n}\left[\left\{\overline{\Upsilon}_{n,d}^{\mathrm{cap}}(O_{1:n},B_{1:n})-\Psi(\mathbb{P})\right\}^2\right] \le E_{\mathbb{P}^{\otimes n}\otimes\mathbb{B}_n\otimes\mathrm{Pois}_{n/4}}\left[\left\{\Upsilon_{n,d}^{\mathrm{cap}}(O_{1:n},B_{1:n},M)-\Psi(\mathbb{P})\right\}^2\right].$$

On $\{M\le n\}$, the prefix $(O_1,B_1),\dots,(O_M,B_M)$ has the corresponding uncapped finite-Poisson marked law restricted to prefixes of length at most n, so the nonoverflow contribution is bounded by

$$E\left[\left\{\widetilde{V}^{\mathrm{unc}}-\Psi(\mathbb{P})\right\}^2\right].$$

On $\{M>n\}$, the capped statistic equals 0, and $\Psi(\mathbb{P})\in[0,1]$, so the overflow contribution is at most

$$\Pr\{\mathrm{Pois}(n/4)>n\}.$$

Thus

$$\mathrm{Risk}_{\mathbb{P}}(\widehat{V}_{n,d}^{\mathrm{JF}},\Psi) \le E\left[\left\{\widetilde{V}^{\mathrm{unc}}-\Psi(\mathbb{P})\right\}^2\right]+\Pr\{\mathrm{Pois}(n/4)>n\}.$$

To bound the Poisson overflow term, let

$$\mathcal{N}_{\mathrm{ov}} \sim \mathrm{Pois}(n/4), \qquad \lambda_{\mathrm{ov}} := \frac{n}{4}, \qquad \gamma_{\mathrm{ov}} := \frac{n}{8}.$$

Since $n \geq 1$, both λ_{ov} and γ_{ov} are positive. For every nonnegative Chernoff tilt ϖ_{ov}, the centered Poisson moment-generating function gives

$$E \exp\{\varpi_{\mathrm{ov}}(\mathcal{N}_{\mathrm{ov}} - \lambda_{\mathrm{ov}})\} = \exp\{\lambda_{\mathrm{ov}}(e^{\varpi_{\mathrm{ov}}} - 1 - \varpi_{\mathrm{ov}})\}.$$

Consequently, for every $\Delta_{\mathrm{ov}} > 0$, Markov's inequality yields

$$\Pr\{\mathcal{N}_{\mathrm{ov}} - \lambda_{\mathrm{ov}} \geq \Delta_{\mathrm{ov}}\} \leq \inf_{\varpi_{\mathrm{ov}} \geq 0} \exp\{-\varpi_{\mathrm{ov}}\Delta_{\mathrm{ov}} + \lambda_{\mathrm{ov}}(e^{\varpi_{\mathrm{ov}}} - 1 - \varpi_{\mathrm{ov}})\}.$$

Taking $\varpi_{\mathrm{ov}} = \log(1 + \Delta_{\mathrm{ov}}/\lambda_{\mathrm{ov}})$ gives the Bennett form

$$\Pr\{\mathcal{N}_{\mathrm{ov}} - \lambda_{\mathrm{ov}} \geq \Delta_{\mathrm{ov}}\} \leq \exp\left[-\left\{(\lambda_{\mathrm{ov}} + \Delta_{\mathrm{ov}})\log\left(1 + \frac{\Delta_{\mathrm{ov}}}{\lambda_{\mathrm{ov}}}\right) - \Delta_{\mathrm{ov}}\right\}\right].$$

The elementary inequality

$$\log(1 + \rho_{\mathrm{ov}}) \geq \frac{2\rho_{\mathrm{ov}}}{\rho_{\mathrm{ov}} + 2}, \qquad \rho_{\mathrm{ov}} \geq 0,$$

implies

$$(\lambda_{\mathrm{ov}} + \Delta_{\mathrm{ov}})\log\left(1 + \frac{\Delta_{\mathrm{ov}}}{\lambda_{\mathrm{ov}}}\right) - \Delta_{\mathrm{ov}} \geq \frac{\Delta_{\mathrm{ov}}^2}{2\lambda_{\mathrm{ov}} + \Delta_{\mathrm{ov}}}.$$

Now choose

$$\Delta_{\mathrm{ov}} := \sqrt{2\lambda_{\mathrm{ov}}\gamma_{\mathrm{ov}}} + 2\gamma_{\mathrm{ov}}.$$

Then

$$\Delta_{\mathrm{ov}}^2 - \gamma_{\mathrm{ov}}(2\lambda_{\mathrm{ov}} + \Delta_{\mathrm{ov}}) = 3\gamma_{\mathrm{ov}}\sqrt{2\lambda_{\mathrm{ov}}\gamma_{\mathrm{ov}}} + 2\gamma_{\mathrm{ov}}^2 \geq 0,$$

so

$$\Pr\{\mathcal{N}_{\mathrm{ov}} - \lambda_{\mathrm{ov}} > \Delta_{\mathrm{ov}}\} \leq e^{-\gamma_{\mathrm{ov}}}.$$

For the present values,

$$\Delta_{\mathrm{ov}} = \sqrt{2(n/4)(n/8)} + 2(n/8) = \frac{n}{2}, \qquad n - \lambda_{\mathrm{ov}} = \frac{3n}{4},$$

and therefore

$$\{\mathcal{N}_{\mathrm{ov}} > n\} \subseteq \{\mathcal{N}_{\mathrm{ov}} - \lambda_{\mathrm{ov}} > \Delta_{\mathrm{ov}}\}.$$

Thus

$$\Pr\{\mathrm{Pois}(n/4) > n\} \leq e^{-n/8}.$$

Finally, $\gamma_{\mathrm{ov}} \leq e^{\gamma_{\mathrm{ov}}}$ follows from $1 + \gamma_{\mathrm{ov}} \leq e^{\gamma_{\mathrm{ov}}}$, and hence

$$\Pr\{\mathrm{Pois}(n/4) > n\} \leq e^{-n/8} \leq \frac{1}{n/8} = \frac{8}{n}.$$

Using $L_d \leq d$,

$$\frac{8}{n} \leq 8\frac{d}{nL_d} = 8\chi_{n,d}.$$

Combining this overflow bound with the uncapped bound from the previous step,

$$\mathrm{Risk}_{\mathbb{P}}(\widehat{V}^{\mathrm{JF}}_{n,d}, \Psi) \leq \left[10000000\{1 + C^{\mathrm{cell}}_{\epsilon} + (C^{\mathrm{cell}}_{\epsilon})^2\} + 8\right] \chi_{n,d} \leq C_{\epsilon}\chi_{n,d}.$$

In the present branch $\chi_{n,d} \leq 1$, hence

$$C_{\epsilon}\chi_{n,d} = C_{\epsilon} \min\left\{1, \frac{d}{n\log(ed)}\right\}.$$

The three branches cover every $n \geq 1$, $d \geq 2$, and $\mathbb{P} \in \mathcal{D}^{\mathrm{obs}}_{d,\epsilon}$, giving the stated uniform bound.

□

Proof of Proposition 3. Fix $(\mathbf{P}, \mathbf{Q}) \in \Delta_d^2$, and put

$$\mathbb{P} = \mathrm{Obs}\{\mathrm{L1Emb}(\mathbf{P}, \mathbf{Q})\}.$$

The sampling condition in Assumption 1 supplies the product-sampling convention used in Definition 5.

1. For $x \in [d]$, define

$$q_{ay,x} = \mathbb{P}\{X = x, A = a, Y = y\}, \qquad a, y \in \{0, 1\}.$$

By Definitions 2 and 16,

$$q_{11,x} = \frac{P_x}{4}, \qquad q_{00,x} = \frac{P_x}{4}, \qquad q_{10,x} = \frac{Q_x}{4}, \qquad q_{01,x} = \frac{Q_x}{4}. \tag{4}$$

Hence

$$p_x = \sum_{a,y\in\{0,1\}} q_{ay,x} = \frac{P_x + Q_x}{2}.$$

The vector $p = (p_x)_{x\in[d]}$ lies in Δ_d, since $p_x \geq 0$ and

$$\sum_{x\in[d]} p_x = \frac{1}{2}\sum_{x\in[d]} P_x + \frac{1}{2}\sum_{x\in[d]} Q_x = 1.$$

Set $\pi_x = 1/2$ for every cell, and set

$$\mu_{1x} = \begin{cases} P_x/(P_x+Q_x), & P_x + Q_x > 0, \\ 0, & P_x + Q_x = 0, \end{cases} \qquad \mu_{0x} = \begin{cases} Q_x/(P_x+Q_x), & P_x + Q_x > 0, \\ 0, & P_x + Q_x = 0. \end{cases}$$

On cells with $P_x + Q_x > 0$, these quantities lie in $[0, 1]$, and the four identities in Equation (4) are exactly the factorization required in Definition 3. On cells with $P_x + Q_x = 0$, nonnegativity gives $P_x = Q_x = 0$, so all four atom masses vanish and the same factorization holds. Moreover,

$$\sum_{y\in\{0,1\}} q_{1y,x} = \sum_{y\in\{0,1\}} q_{0y,x} = \frac{P_x + Q_x}{4},$$

so whenever $p_x > 0$,

$$\pi_x = \frac{\sum_y q_{1y,x}}{p_x} = \frac{(P_x+Q_x)/4}{(P_x+Q_x)/2} = \frac{1}{2}.$$

Because $0 < \epsilon < 1/2$, this propensity lies in $[\epsilon, 1-\epsilon]$. Thus $\mathbb{P} \in \mathcal{D}^{\mathrm{obs}}_{d,\epsilon}$.

2. Write

$$\|\mathbf{P}-\mathbf{Q}\|_1 := \sum_{x\in[d]} |P_x - Q_x|.$$

By Definition 4,

$$\Psi(\mathbb{P}) = \sum_{x\in[d]} p_x \max\{\mu_{0x}, \mu_{1x}\}.$$

A null cell contributes zero. On a cell with $P_x + Q_x > 0$,

$$p_x \max\{\mu_{0x}, \mu_{1x}\} = \frac{P_x + Q_x}{2} \max\left\{\frac{Q_x}{P_x+Q_x}, \frac{P_x}{P_x+Q_x}\right\} = \frac{\max\{P_x, Q_x\}}{2}.$$

For each x,

$$\frac{\max\{P_x, Q_x\}}{2} = \frac{P_x+Q_x}{4} + \frac{|P_x - Q_x|}{4}.$$

Summing over x and using $\sum_x P_x = \sum_x Q_x = 1$ gives

$$\Psi(\mathbb{P}) = \frac{1}{2} + \frac{1}{4}\|\mathbf{P}-\mathbf{Q}\|_1 = \frac{1}{2} + \frac{1}{4}\sum_{x\in[d]} |P_x - Q_x|. \tag{5}$$

3. Define the single-observation kernel K_1 from a paired draw $\eta = (\eta_{\mathrm{P}}, \eta_{\mathrm{Q}}) \in [d]\times[d]$ by

$$K_1\big(\eta, \{(x,a,y)\}\big) = \begin{cases} 1/4, & (x,a,y) \in \{(\eta_{\mathrm{P}},0,0), (\eta_{\mathrm{P}},1,1), (\eta_{\mathrm{Q}},0,1), (\eta_{\mathrm{Q}},1,0)\}, \\ 0, & \text{otherwise.} \end{cases}$$

Under one paired draw from $\mathbf{P}\otimes\mathbf{Q}$,

$$\Pr\{\eta = (x_{\mathrm{P}}, x_{\mathrm{Q}})\} = P_{x_{\mathrm{P}}} Q_{x_{\mathrm{Q}}}.$$

Therefore

$$\{(\mathbf{P}\otimes\mathbf{Q})K_1\}\{(x,1,1)\} = \sum_{x_{\mathrm{Q}}\in[d]} \frac{1}{4} P_x Q_{x_{\mathrm{Q}}} = \frac{P_x}{4},$$

and similarly

$$\{(\mathbf{P}\otimes\mathbf{Q})K_1\}\{(x,0,0)\} = \frac{P_x}{4}, \qquad \{(\mathbf{P}\otimes\mathbf{Q})K_1\}\{(x,1,0)\} = \frac{Q_x}{4}, \qquad \{(\mathbf{P}\otimes\mathbf{Q})K_1\}\{(x,0,1)\} = \frac{Q_x}{4}.$$

These are the atom masses in Equation (4), so the one-observation kernelized paired draw has law $\mathbb{P}$.

4. Define the fixed-n product kernel K_n on $([d]\times[d])^n$ by

$$K_n\big(\xi, B_1\times\cdots\times B_n\big) = \prod_{i=1}^{n} K_1(\xi_i, B_i), \qquad \xi = (\xi_1, \ldots, \xi_n).$$

For $n = 0$, this is the empty product kernel; for $n > 0$, it applies K_1 independently in each coordinate. Since the fixed paired-sample law is $(\mathbf{P}\otimes\mathbf{Q})^{\otimes n}$ and the one-coordinate kernelized law is $\mathbb{P}$, finite-product independence gives

$$K_n\{(\mathbf{P}\otimes\mathbf{Q})^{\otimes n}\} = \mathbb{P}^{\otimes n}. \tag{6}$$

This is the explicit fixed-n product Markov kernel asserted in the proposition.

5. Let $\widehat{V}$ be any measurable observed-sample estimator. Since the observed sample space is finite, $\widehat{V}$ is bounded. Define the paired-sample pullback

$$\widehat{L}_{\widehat{V}}(\xi) = 4\left\{\int \widehat{V}(o)\, K_n(\xi, do) - \frac{1}{2}\right\}, \qquad \xi \in ([d] \times [d])^n.$$

For each fixed paired sample ξ, Equation (5) and Jensen's inequality for $t \mapsto t^2$ give

$$\frac{1}{16}\left\{\widehat{L}_{\widehat{V}}(\xi) - \|\mathbf{P}-\mathbf{Q}\|_1\right\}^2 = \left[\int \{\widehat{V}(o) - \Psi(\mathbb{P})\}\, K_n(\xi, do)\right]^2 \leq \int \{\widehat{V}(o) - \Psi(\mathbb{P})\}^2\, K_n(\xi, do).$$

Integrating under $(\mathbf{P}\otimes\mathbf{Q})^{\otimes n}$ and using Equation (6) yields

$$\frac{1}{16} E_{(\mathbf{P}\otimes\mathbf{Q})^{\otimes n}}\left[\left\{\widehat{L}_{\widehat{V}} - \|\mathbf{P}-\mathbf{Q}\|_1\right\}^2\right] \leq E_{\mathbb{P}^{\otimes n}}\left[\{\widehat{V} - \Psi(\mathbb{P})\}^2\right].$$

Taking the supremum over $(\mathbf{P},\mathbf{Q}) \in \Delta_d^2$, and using the admissibility proved above, gives

$$\sup_{(\mathbf{P},\mathbf{Q})\in\Delta_d^2} E_{(\mathbf{P}\otimes\mathbf{Q})^{\otimes n}}\left[\left\{\widehat{L}_{\widehat{V}} - \|\mathbf{P}-\mathbf{Q}\|_1\right\}^2\right] \leq 16 \sup_{\widetilde{\mathbb{P}}\in\mathcal{D}_{d,\epsilon}^{\mathrm{obs}}} E_{\widetilde{\mathbb{P}}^{\otimes n}}\left[\{\widehat{V} - \Psi(\widetilde{\mathbb{P}})\}^2\right].$$

The estimator $\widehat{L}_{\widehat{V}}$ is admissible for the paired problem in Definition 15; the zero estimator and the fixed pair chosen at the start give the required nonempty estimator and parameter classes. Hence

$$\mathfrak{R}_{n,d}^{L_1} \leq 16 \sup_{\widetilde{\mathbb{P}}\in\mathcal{D}_{d,\epsilon}^{\mathrm{obs}}} E_{\widetilde{\mathbb{P}}^{\otimes n}}\left[\{\widehat{V} - \Psi(\widetilde{\mathbb{P}})\}^2\right].$$

Taking the infimum over all observed-sample estimators $\widehat{V}$, as in Definition 5, gives

$$\frac{1}{16}\mathfrak{R}_{n,d}^{L_1} \leq \mathfrak{R}_{n,d,\epsilon}.$$

6. Combining the observed-model membership, the displayed formulas for p_x, π_x, and $\Psi(\mathbb{P})$, the product-kernel identity, and the minimax transfer proves every asserted conclusion for the fixed pair $(\mathbf{P},\mathbf{Q})$. Since the pair was arbitrary, the proposition follows.

□

Proof of Theorem 2. Fix $0 < \epsilon < 1/2$. For an alphabet size $r \geq 2$, set

$$L_r = \log(er), \qquad L_n^\circ = \log(en),$$

where L_n° is used only for sample-size logarithms. Let $\mathfrak{R}_{n,r}^{L_1}$ be the fixed-sample two-sample L_1 risk of Definition 15. Also define the Poissonized two-sample risk

$$\mathfrak{R}_{n,r}^{L_1,\mathrm{Pois}} := \inf_{\widehat{D}} \sup_{(\mathbf{P},\mathbf{Q})\in\Delta_r^2} E_{\mathcal{H}_{n,\mathbf{P},\mathbf{Q}}}\left[\left\{\widehat{D}(N^{\mathbf{P}}, N^{\mathbf{Q}}) - \|\mathbf{P}-\mathbf{Q}\|_1\right\}^2\right],$$

where, under $\mathcal{H}_{n,\mathbf{P},\mathbf{Q}}$, the two histograms have independent coordinates

$$N_x^{\mathbf{P}} \sim \mathrm{Pois}(2nP_x), \qquad N_x^{\mathbf{Q}} \sim \mathrm{Pois}(2nQ_x), \qquad x \in [r],$$

and the infimum ranges over measurable histogram statistics whose squared loss is integrable for every $(\mathbf{P},\mathbf{Q}) \in \Delta_r^2$ and whose displayed worst-case risk is finite.

1. Lemma 10, using the moment-matching priors of [Cai and Low, 2011, Lemma 1 and Section 3.1], supplies an integer $d_{\text{dense}} \geq 2$ and a constant $\gamma_{\text{dense},\epsilon} > 0$ such that

$$r \geq d_{\text{dense}}, \qquad r^2 < n \quad \Longrightarrow \quad \gamma_{\text{dense},\epsilon} \frac{r}{nL_r} \leq \mathfrak{R}_{n,r,\epsilon}.$$

The Poissonized lower bound of [Jiao et al., 2018, Theorem 3, equation (24), and Section 3.1], with constants 1/4 and 2, gives a universal $\gamma_{\text{JHW}} > 0$ such that

$$r \geq 2, \qquad \frac{r}{4L_r} \leq n, \qquad L_n^\circ \leq 2L_r \quad \Longrightarrow \quad \gamma_{\text{JHW}} \min\left\{1, \frac{r}{nL_n^\circ}\right\} \leq \mathfrak{R}_{n,r}^{L_1,\text{Pois}}.$$

For fixed-sample estimators this yields

$$\gamma_{\text{JHW}} \min\left\{1, \frac{r}{nL_n^\circ}\right\} - 8\exp\{-n(1-\log 2)\} \leq \mathfrak{R}_{n,r}^{L_1}.$$

Indeed, from any fixed-sample estimator $\widehat{D}$, construct a histogram statistic as follows. Let

$$\mathcal{E}_n^{\text{pois}} = \left\{\sum_{x=1}^{r} N_x^{\mathbf{P}} \geq n, \ \sum_{x=1}^{r} N_x^{\mathbf{Q}} \geq n\right\}.$$

On $\mathcal{E}_n^{\text{pois}}$, take the conditional expectation of $\widehat{D}$ applied to ordered n-samples drawn from the two histograms; on $(\mathcal{E}_n^{\text{pois}})^c$, output 0. On $\mathcal{E}_n^{\text{pois}}$, the conditional reconstruction has the fixed n-sample law, so Jensen's inequality gives the fixed-sample risk. Since $\|\mathbf{P} - \mathbf{Q}\|_1 \in [0, 2]$, the squared loss on $(\mathcal{E}_n^{\text{pois}})^c$ is at most 4. The Chernoff bound for $\text{Pois}(2n)$ gives

$$\Pr\left\{\sum_x N_x^{\mathbf{P}} < n\right\} = \Pr\left\{\sum_x N_x^{\mathbf{Q}} < n\right\} \leq \exp\{-n(1-\log 2)\},$$

and hence the displayed fixed-sample lower bound follows after taking the infimum over $\widehat{D}$.

Choose $d_{\log}$ so that

$$r \geq d_{\log} \quad \Longrightarrow \quad L_r \leq \frac{r}{2},$$

and choose $n_{\text{tail}} \geq 2$ so that, for every $n \geq n_{\text{tail}}$,

$$8\exp\{-n(1-\log 2)\} \leq \frac{\gamma_{\text{JHW}}}{2n}, \qquad 8\exp\{-n(1-\log 2)\} \leq \frac{\gamma_{\text{JHW}}}{2}.$$

Set

$$d_\star = \max\{d_{\text{dense}}, d_{\log}\}, \qquad \Xi_\star = \max\{d_\star, n_{\text{tail}}\},$$

and

$$c_\epsilon = \min\left\{\gamma_{\text{dense},\epsilon}, \frac{\gamma_{\text{JHW}}}{64}, \frac{1}{100\Xi_\star}\right\} > 0.$$

2. For every $n \geq 1$ and $d \geq 2$,

$$\frac{1}{100n} \leq \mathfrak{R}_{n,d,\epsilon}.$$

Put

$$\delta_n = \frac{2}{5\sqrt{n}}.$$

In every cell $x \in [d]$, take $p_x = 1/d$, $\pi_x = 1/2$, and $\mu_{0x} = 1/2$. Under $\mathbb{P}^{\mathrm{reg}}_{0,n,d}$, set $\mu_{1x} = 1/2$; under $\mathbb{P}^{\mathrm{reg}}_{1,n,d}$, set $\mu_{1x} = 1/2 + \delta_n$. Since $d \geq 2$, the vector with coordinates $p_x = 1/d$ lies in Δ_d. The fixed range $0 < \epsilon < 1/2$ gives $1/2 \in [\epsilon, 1-\epsilon]$, so the displayed propensities satisfy Assumption 4. Finally $0 \leq \delta_n \leq 2/5$ places all arm-specific means in $[0, 1]$. Thus both laws belong to $\mathcal{D}^{\mathrm{obs}}_{d,\epsilon}$, and

$$\Psi(\mathbb{P}^{\mathrm{reg}}_{0,n,d}) = \frac{1}{2}, \qquad \Psi(\mathbb{P}^{\mathrm{reg}}_{1,n,d}) = \frac{1}{2} + \delta_n.$$

The one-observation chi-square divergence is

$$\chi^2(\mathbb{P}^{\mathrm{reg}}_{1,n,d}, \mathbb{P}^{\mathrm{reg}}_{0,n,d}) = \frac{(1/2)\delta_n^2}{(1/2)(1-1/2)} = 2\delta_n^2,$$

so tensorization and $1 + t \leq e^t$ give

$$1 + \chi^2\Big((\mathbb{P}^{\mathrm{reg}}_{1,n,d})^{\otimes n}, (\mathbb{P}^{\mathrm{reg}}_{0,n,d})^{\otimes n}\Big) = (1 + 2\delta_n^2)^n \leq \exp(2n\delta_n^2) \leq 2,$$

because $2n\delta_n^2 = 8/25 \leq \log 2$. Therefore

$$\mathrm{TV}\Big((\mathbb{P}^{\mathrm{reg}}_{0,n,d})^{\otimes n}, (\mathbb{P}^{\mathrm{reg}}_{1,n,d})^{\otimes n}\Big) \leq \frac{1}{2}\sqrt{\chi^2\Big((\mathbb{P}^{\mathrm{reg}}_{1,n,d})^{\otimes n}, (\mathbb{P}^{\mathrm{reg}}_{0,n,d})^{\otimes n}\Big)} \leq \frac{1}{2}.$$

For two model points with target separation $2s$ and product-law total variation at most $1/2$, thresholding any estimator at the target midpoint and using

$$E_j\{(\widehat{V} - \Psi_j)^2\} \geq s^2 \Pr_j\{|\widehat{V} - \Psi_j| \geq s\}, \qquad j \in \{0, 1\},$$

gives a worst-case risk at least $s^2/4$. With $s = \delta_n/2$,

$$\mathfrak{R}_{n,d,\epsilon} \geq \frac{(\delta_n/2)^2}{4} = \frac{1}{100n}.$$

3. If $n < n_{\mathrm{tail}}$, then $n \leq \Xi_\star$, and therefore

$$c_\epsilon \min\left\{1, \frac{d}{nL_d}\right\} \leq c_\epsilon \leq \frac{1}{100\Xi_\star} \leq \frac{1}{100n} \leq \mathfrak{R}_{n,d,\epsilon}.$$

4. Assume $n \geq n_{\mathrm{tail}}$. If $d < d_\star$, then $d \leq \Xi_\star$. Since $L_d \geq 1$,

$$c_\epsilon \min\left\{1, \frac{d}{nL_d}\right\} \leq c_\epsilon \frac{d}{nL_d} \leq \frac{d}{100\Xi_\star nL_d} \leq \frac{1}{100n} \leq \mathfrak{R}_{n,d,\epsilon}.$$

5. Assume $n \geq n_{\mathrm{tail}}$, $d \geq d_\star$, and $d^2 < n$. The dense lower bound from Lemma 10 gives

$$\gamma_{\mathrm{dense},\epsilon} \frac{d}{nL_d} \leq \mathfrak{R}_{n,d,\epsilon}.$$

Since $c_\epsilon \leq \gamma_{\mathrm{dense},\epsilon}$ and $\min\{1, d/(nL_d)\} \leq d/(nL_d)$,

$$c_\epsilon \min\left\{1, \frac{d}{nL_d}\right\} \leq \mathfrak{R}_{n,d,\epsilon}.$$

6. It remains to treat $n \geq n_{\text{tail}}$, $d \geq d_\star$, and $n \leq d^2$. Then

$$L_n^\circ \leq \log(ed^2) = 1 + 2\log d \leq 2(1 + \log d) = 2L_d.$$

Because $d \geq d_{\log}$, also $L_d \leq d/2$, and hence

$$L_n^\circ \leq 2L_d \leq d.$$

7. Suppose first that

$$n < \frac{d}{4L_d}.$$

Define

$$s_n = \lceil nL_n^\circ \rceil .$$

The preceding inequality implies $nL_d < d/4$. Since $L_d \geq 1$, this gives $n < d$, hence $L_n^\circ \leq L_d$, and therefore

$$nL_n^\circ \leq nL_d < \frac{d}{4} < d, \qquad 2 \leq s_n \leq d.$$

The ceiling definition also yields the lower bound

$$nL_n^\circ \leq s_n.$$

Because $L_n^\circ \geq 1$ and $n \geq n_{\text{tail}} \geq 2$, this implies $n \leq s_n$, and monotonicity of the logarithm gives

$$L_n^\circ = \log(en) \leq \log(es_n).$$

The upper ceiling estimate is just as important: since $nL_n^\circ \geq 2$,

$$s_n = \lceil nL_n^\circ \rceil < nL_n^\circ + 1 \leq 2nL_n^\circ.$$

The denominator below is positive because $s_n \geq 2$. Consequently,

$$\frac{s_n}{4\log(es_n)} \leq \frac{2nL_n^\circ}{4\log(es_n)} \leq \frac{n}{2} \leq n,$$

and the two remaining gate estimates are

$$L_n^\circ \leq 2\log(es_n), \qquad 1 \leq \frac{s_n}{nL_n^\circ},$$

the second following by dividing $nL_n^\circ \leq s_n$ by the positive product nL_n°. The fixed-sample L_1 bound at alphabet s_n yields

$$\gamma_{\text{JHW}} - 8\exp\{-n(1 - \log 2)\} \leq \mathfrak{R}_{n,s_n}^{L_1},$$

and the choice of n_{tail} gives

$$\frac{\gamma_{\text{JHW}}}{2} \leq \mathfrak{R}_{n,s_n}^{L_1}.$$

Embedding Δ_{s_n} into Δ_d by padding the remaining coordinates with zero preserves the L_1 distance, so

$$\mathfrak{R}_{n,s_n}^{L_1} \leq \mathfrak{R}_{n,d}^{L_1}.$$

By Proposition 3,

$$\frac{1}{16}\mathfrak{R}^{L_1}_{n,d} \le \mathfrak{R}_{n,d,\epsilon},$$

and consequently

$$\frac{\gamma_{\mathrm{JHW}}}{32} \le \mathfrak{R}_{n,d,\epsilon}.$$

The saturated inequality gives $d/(nL_d) > 4$, hence

$$\min\left\{1, \frac{d}{nL_d}\right\} = 1.$$

Since $c_\epsilon \le \gamma_{\mathrm{JHW}}/64$,

$$c_\epsilon \min\left\{1, \frac{d}{nL_d}\right\} \le \mathfrak{R}_{n,d,\epsilon}.$$

8. Finally suppose that

$$\frac{d}{4L_d} \le n.$$

The fixed-sample L_1 bound at alphabet d gives

$$\gamma_{\mathrm{JHW}} \min\left\{1, \frac{d}{nL_n^\circ}\right\} - 8\exp\{-n(1-\log 2)\} \le \mathfrak{R}^{L_1}_{n,d}.$$

Since $L_n^\circ \le d$,

$$\frac{1}{n} \le \min\left\{1, \frac{d}{nL_n^\circ}\right\}.$$

The choice of n_{tail} therefore yields

$$\frac{\gamma_{\mathrm{JHW}}}{2} \min\left\{1, \frac{d}{nL_n^\circ}\right\} \le \mathfrak{R}^{L_1}_{n,d}.$$

Also $L_n^\circ \le 2L_d$, so

$$\frac{d}{nL_d} \le 2\frac{d}{nL_n^\circ}, \qquad \min\left\{1, \frac{d}{nL_d}\right\} \le 2\min\left\{1, \frac{d}{nL_n^\circ}\right\}.$$

Using $c_\epsilon \le \gamma_{\mathrm{JHW}}/64$ and Proposition 3,

$$c_\epsilon \min\left\{1, \frac{d}{nL_d}\right\} \le \frac{\gamma_{\mathrm{JHW}}}{32} \min\left\{1, \frac{d}{nL_n^\circ}\right\} \le \frac{\mathfrak{R}^{L_1}_{n,d}}{16} \le \mathfrak{R}_{n,d,\epsilon}.$$

The displayed alternatives cover every $n \ge 1$ and $d \ge 2$. Thus the constant $c_\epsilon > 0$ satisfies

$$c_\epsilon \min\left\{1, \frac{d}{n\log(ed)}\right\} \le \mathfrak{R}_{n,d,\epsilon}$$

for all $n \ge 1$ and $d \ge 2$. $\square$

Proof of Theorem 3. 1. Assumption 1 identifies the sampling law with the canonical product law. Apply Theorem 1 with this sampling law and choose its universal Jackson tuning constants $H_0 > 0$, $\kappa \in (0,1)$, and cutoff $D_0 > 2$. Fix $\epsilon \in (0, 1/2)$. The same result gives a constant $C_\epsilon^{\mathrm{JF}} > 0$ such that, for all $n \geq 1$, $d \geq 2$, and $\mathbb{P} \in \mathcal{D}_{d,\epsilon}^{\mathrm{obs}}$,

$$E_{\mathbb{P}^{\otimes n}}\left[\left(\widehat{V}_{n,d}^{\mathrm{JF}} - \Psi(\mathbb{P})\right)^2\right] \leq C_\epsilon^{\mathrm{JF}} \min\left\{1, \frac{d}{n\log(ed)}\right\}.$$

The lower-bound input is Theorem 2, using the moment-matching priors of [Cai and Low, 2011, Lemma 1 and Section 3.1] and the Poissonized L_1 minimax lower bound of [Jiao et al., 2018, Theorem 3 and equation (24)]. It gives a constant $c_\epsilon^{\mathrm{low}} > 0$ such that, for all $n \geq 1$ and $d \geq 2$,

$$c_\epsilon^{\mathrm{low}} \min\left\{1, \frac{d}{n\log(ed)}\right\} \leq \mathfrak{R}_{n,d,\epsilon}.$$

Set

$$c_\epsilon := c_\epsilon^{\mathrm{low}}, \qquad C_\epsilon := C_\epsilon^{\mathrm{JF}} + c_\epsilon^{\mathrm{low}}.$$

Then $0 < c_\epsilon \leq C_\epsilon$. For later comparison, when $n \geq 1$ and $d \geq 2$, $L_d = \log(ed) > 0$, and hence

$$\rho_{n,d} := \min\left\{1, \frac{d}{nL_d}\right\} \geq 0.$$

2. The preceding lower bound gives

$$c_\epsilon \rho_{n,d} \leq \mathfrak{R}_{n,d,\epsilon}.$$

For the middle inequality, define for each measurable observed-data estimator $\widehat{V}$

$$\mathcal{E}_{n,d,\epsilon}(\widehat{V}) := \sup_{\mathbb{P} \in \mathcal{D}_{d,\epsilon}^{\mathrm{obs}}} E_{\mathbb{P}^{\otimes n}}\left[\left(\widehat{V} - \Psi(\mathbb{P})\right)^2\right].$$

By Definition 5,

$$\mathfrak{R}_{n,d,\epsilon} = \inf_{\widehat{V}} \mathcal{E}_{n,d,\epsilon}(\widehat{V}),$$

where the infimum is over the same measurable sample maps. The estimator $\widehat{V}_{n,d}^{\mathrm{JF}}$ in Algorithm 1 is one admissible map, and squared-error risks are nonnegative:

$$E_{\mathbb{P}^{\otimes n}}\left[\left(\widehat{V}_{n,d}^{\mathrm{JF}} - \Psi(\mathbb{P})\right)^2\right] \geq 0.$$

Therefore evaluating the infimum at this estimator gives

$$\mathfrak{R}_{n,d,\epsilon} \leq \mathcal{E}_{n,d,\epsilon}(\widehat{V}_{n,d}^{\mathrm{JF}}).$$

3. It remains to bound this particular worst-case risk. If the set over which the displayed supremum is taken is empty, the real-supremum convention gives

$$\mathcal{E}_{n,d,\epsilon}(\widehat{V}_{n,d}^{\mathrm{JF}}) = 0,$$

and $0 \le C_\epsilon \rho_{n,d}$. Otherwise, for every $\mathbb{P} \in \mathcal{D}^{\mathrm{obs}}_{d,\epsilon}$, the upper bound from Theorem 1 gives

$$E_{\mathbb{P}^{\otimes n}}\left[\left(\widehat{V}^{\mathrm{JF}}_{n,d} - \Psi(\mathbb{P})\right)^2\right] \le C^{\mathrm{JF}}_\epsilon \rho_{n,d} \le (C^{\mathrm{JF}}_\epsilon + c^{\mathrm{low}}_\epsilon)\rho_{n,d} = C_\epsilon \rho_{n,d},$$

where the second inequality uses $\rho_{n,d} \ge 0$. Taking the supremum over $\mathbb{P}$ yields

$$\mathcal{E}_{n,d,\epsilon}(\widehat{V}^{\mathrm{JF}}_{n,d}) \le C_\epsilon \min\left\{1, \frac{d}{nL_d}\right\}.$$

4. Finally fix a measurable observed-data estimator $\widehat{V}$ and define its causal worst-case risk by

$$\mathcal{E}^{\mathrm{causal}}_{n,d,\epsilon}(\widehat{V}) := \sup_{P \in \mathcal{P}_{d,\epsilon}} E_{\mathrm{Obs}(P)^{\otimes n}}\left[\left(\widehat{V} - V^\star(P)\right)^2\right].$$

The two attained-risk sets are equal. Indeed, if $P \in \mathcal{P}_{d,\epsilon}$, then Proposition 1 gives

$$\mathrm{Obs}(P) \in \mathcal{D}^{\mathrm{obs}}_{d,\epsilon},$$

and Proposition 2 gives

$$V^\star(P) = \Psi\{\mathrm{Obs}(P)\}.$$

Thus each causal risk value is an observed-law risk value:

$$E_{\mathrm{Obs}(P)^{\otimes n}}\left[\left(\widehat{V} - V^\star(P)\right)^2\right] = E_{\mathrm{Obs}(P)^{\otimes n}}\left[\left(\widehat{V} - \Psi\{\mathrm{Obs}(P)\}\right)^2\right].$$

Conversely, if $\mathbb{P} \in \mathcal{D}^{\mathrm{obs}}_{d,\epsilon}$, then Proposition 1 supplies $P \in \mathcal{P}_{d,\epsilon}$ with

$$\mathrm{Obs}(P) = \mathbb{P},$$

and Proposition 2 again gives

$$V^\star(P) = \Psi(\mathbb{P}).$$

Hence the corresponding observed-law risk is attained in the causal class:

$$E_{\mathbb{P}^{\otimes n}}\left[\left(\widehat{V} - \Psi(\mathbb{P})\right)^2\right] = E_{\mathrm{Obs}(P)^{\otimes n}}\left[\left(\widehat{V} - V^\star(P)\right)^2\right].$$

Therefore

$$\mathcal{E}^{\mathrm{causal}}_{n,d,\epsilon}(\widehat{V}) = \mathcal{E}_{n,d,\epsilon}(\widehat{V})$$

for every measurable observed-data estimator $\widehat{V}$. Taking the infimum over the common estimator class in Definitions 5 and 17 gives

$$\mathfrak{R}^{\mathrm{causal}}_{n,d,\epsilon} = \mathfrak{R}_{n,d,\epsilon}.$$

Together with the three inequalities above, this proves the result.

□

Proof of Theorem 4. Fix $\epsilon \in (0, 1/2)$. The published moment-matching conclusion of [Cai and Low, 2011, Lemma 1 and Section 3.1] and the published Poissonized L_1 lower bound of [Jiao et al., 2018, Theorem 3, equation (24)] enter through Theorem 3 and Proposition 4. By Theorem 3, there are constants c_ϵ, C_ϵ with $0 < c_\epsilon \le C_\epsilon$ such that, for every $n \ge 1$ and $d \ge 2$,

$$c_\epsilon \Phi_{n,d} \le \mathfrak{R}_{n,d,\epsilon} \le C_\epsilon \Phi_{n,d}, \qquad \mathfrak{R}^{\mathrm{causal}}_{n,d,\epsilon} = \mathfrak{R}_{n,d,\epsilon},$$

where

$$\Phi_{n,d} := \min\left\{1, \frac{d}{n \log(ed)}\right\}.$$

The upper comparison uses the Jackson-factorial worst-case risk term displayed in Theorem 3.

Let $d_\bullet = (d_n)_{n \ge 1}$ satisfy $d_n \ge 2$ for every n. Define

$$\Phi_n := \Phi_{n,d_n}, \qquad \rho_n := \frac{d_n}{n \log(ed_n)}, \qquad \chi_n := \frac{d_n}{n \log(en)}.$$

All displayed denominators are positive. We prove

$$\Phi_n \to 0 \quad \Longleftrightarrow \quad \chi_n \to 0,$$

which is exactly $d_n = o\{n \log(en)\}$. Suppose first that $\Phi_n \to 0$. Then eventually $\Phi_n < 1/2$, hence $\Phi_n = \rho_n$, so $\rho_n \to 0$. For all sufficiently large n,

$$0 \le \chi_n \le \frac{1}{\log(en)} + 2\rho_n.$$

Indeed, if $d_n < n$, then

$$\chi_n = \frac{d_n/n}{\log(en)} \le \frac{1}{\log(en)}.$$

If $n \le d_n$, set

$$\tau_n := \frac{d_n}{n} \ge 1.$$

Then

$$\log(ed_n) = \log(en) + \log \tau_n, \qquad \rho_n = \frac{\tau_n}{\log(en) + \log \tau_n}.$$

Since eventually $\rho_n \le 1/2$ and $\log \tau_n \le \tau_n$,

$$2\tau_n \le \log(en) + \log \tau_n \le \log(en) + \tau_n,$$

so $\tau_n \le \log(en)$. Consequently,

$$\log(ed_n) \le 2 \log(en), \qquad \chi_n = \frac{\tau_n}{\log(en)} \le 2\frac{\tau_n}{\log(ed_n)} = 2\rho_n.$$

Because $1/\log(en) \to 0$, this gives $\chi_n \to 0$.

Conversely, suppose $\chi_n \to 0$. For all $n \ge 1$,

$$0 \le \rho_n \le n^{-1/2} + 2\chi_n.$$

If $d_n \le \sqrt{n}$, then $\log(ed_n) \ge 1$, and therefore

$$\rho_n \le \frac{d_n}{n} \le n^{-1/2}.$$

If $\sqrt{n} < d_n$, then $n < d_n^2$, whence

$$en \le (ed_n)^2, \qquad \log(en) \le 2\log(ed_n),$$

and therefore

$$\rho_n \le 2\chi_n.$$

Since $n^{-1/2} \to 0$, we get $\rho_n \to 0$, and then

$$\Phi_n = \min\{1, \rho_n\} \to 0.$$

The two-sided comparison with Φ_n gives

$$\mathfrak{R}_{n,d_n,\epsilon} \to 0 \quad \Longleftrightarrow \quad \Phi_n \to 0.$$

For the forward implication, the lower comparison gives

$$0 \le \Phi_n \le c_\epsilon^{-1}\mathfrak{R}_{n,d_n,\epsilon}$$

for all $n \ge 1$. For the reverse implication, the upper comparison gives

$$0 \le \mathfrak{R}_{n,d_n,\epsilon} \le C_\epsilon \Phi_n$$

for all $n \ge 1$. Combining this equivalence with the preceding display proves observed consistency.

We next prove the observed parametric boundary. Suppose

$$\frac{\zeta_-}{n} \le \mathfrak{R}_{n,d_n,\epsilon} \le \frac{\zeta_+}{n}$$

eventually for constants $0 < \zeta_- \le \zeta_+$. The upper parametric bound and the lower frontier bound imply

$$n\Phi_n \le \frac{\zeta_+}{c_\epsilon}$$

eventually. For all sufficiently large $n > \zeta_+/c_\epsilon$, the alternative $\rho_n > 1$ would give $\Phi_n = 1$ and hence $n\Phi_n = n > \zeta_+/c_\epsilon$, a contradiction. Thus eventually $\rho_n \le 1$, and so

$$\frac{d_n}{\log(ed_n)} = n\rho_n = n\Phi_n \le \frac{\zeta_+}{c_\epsilon}.$$

For $d_n \ge 2$,

$$\log(ed_n) = 1 + \log d_n = 1 + 2\log\sqrt{d_n} \le 1 + 2(\sqrt{d_n} - 1) \le 2\sqrt{d_n}.$$

Therefore

$$\sqrt{d_n} \le 2\frac{d_n}{\log(ed_n)} \le 2\frac{\zeta_+}{c_\epsilon},$$

and hence d_n is eventually bounded. This is $d_n = O(1)$.

Conversely, assume $d_n = O(1)$. Then there are a real number K_{bd} and an index N_{bd} such that $d_n \le K_{\mathrm{bd}}$ for all $n \ge N_{\mathrm{bd}}$. Enlarging over the finite set $1 \le n < N_{\mathrm{bd}}$, choose an integer d_{bd} such that

$$d_n \le d_{\mathrm{bd}} \qquad \text{for every } n \ge 1.$$

This finite enlargement is the passage from an eventual $O(1)$ bound to a single global alphabet bound.

The bounded-alphabet sequence clause of Proposition 4 applied with this d_{bd} supplies constants η_-, η_+ with $0 < \eta_- \le \eta_+$ such that, for all sufficiently large n,

$$\frac{\eta_-}{n} \le \mathfrak{R}_{n,d_n,\epsilon} \le \frac{\eta_+}{n}.$$

Together with the preceding implication, this proves the observed parametric equivalence.

It remains to transfer the two equivalences to the causal risk. The equality supplied by Theorem 3 gives, for every $n \ge 1$,

$$\mathfrak{R}^{\mathrm{causal}}_{n,d_n,\epsilon} = \mathfrak{R}_{n,d_n,\epsilon}.$$

Hence

$$\mathfrak{R}^{\mathrm{causal}}_{n,d_n,\epsilon} \to 0 \quad \Longleftrightarrow \quad \mathfrak{R}_{n,d_n,\epsilon} \to 0,$$

and observed consistency gives

$$\mathfrak{R}^{\mathrm{causal}}_{n,d_n,\epsilon} \to 0 \quad \Longleftrightarrow \quad d_n = o\{n \log(en)\}.$$

The same equality transports the parametric-rate property itself:

$$\mathfrak{R}^{\mathrm{causal}}_{n,d_n,\epsilon} \text{ has the parametric } n^{-1} \text{ rate} \quad \Longleftrightarrow \quad \mathfrak{R}_{n,d_n,\epsilon} \text{ has the parametric } n^{-1} \text{ rate}.$$

Indeed, the same eventual constants bound both sequences after replacing one risk by the other in the displayed equality. Composing this equivalence with the observed parametric equivalence proves the causal parametric boundary. □

Proof of Proposition 4. The published moment-matching construction and absolute-value approximation rate of [Cai and Low, 2011, Lemma 1 and Section 3.1], together with the published Poissonized two-sample L_1 lower-bound argument of [Jiao et al., 2018, Theorem 3 and equation (24)], are the literature inputs used by Theorem 3. Apply Theorem 3 and choose the tuning constants it supplies. Write them as H_0, κ, and D_0. They satisfy

$$H_0 > 0, \qquad 0 < \kappa < 1, \qquad D_0 > 2.$$

For each fixed $\epsilon \in (0, 1/2)$, Theorem 3 gives constants c^{mf}_ϵ and C^{mf}_ϵ, with

$$0 < c^{\mathrm{mf}}_\epsilon \le C^{\mathrm{mf}}_\epsilon,$$

such that, for every $n \ge 1$ and $d \ge 2$, with

$$L_d = \log(ed),$$

one has

$$c^{\mathrm{mf}}_\epsilon \min\left\{1, \frac{d}{nL_d}\right\} \le \mathfrak{R}_{n,d,\epsilon} \le C^{\mathrm{mf}}_\epsilon \min\left\{1, \frac{d}{nL_d}\right\}.$$

1. Fix $n \ge 1$ and $d \ge 2$. The logarithms L_d and $\log(en)$ are positive. Since

$$L_d = 1 + \log d = 1 + 2\log\sqrt{d} \le 1 + 2(\sqrt{d} - 1) \le 2\sqrt{d},$$

we have

$$\frac{\sqrt{d}}{n} \leq 2\,\frac{d}{nL_d}.$$

Define

$$\tau_{n,d} = \frac{d}{nL_d}.$$

We prove

$$\min\left\{1, \frac{\sqrt{d}}{n} + \frac{d}{n\log(en)}\right\} \leq 4\min\left\{1, \tau_{n,d}\right\}.$$

If $\tau_{n,d} \geq 1$, the right-hand side is 4, while the left-hand side is at most 1. Suppose $\tau_{n,d} < 1$. If $d \leq n$, then $ed \leq en$, so

$$L_d \leq \log(en) \leq 2\log(en).$$

If $d > n$, put

$$\eta_{n,d} = \frac{d}{n}.$$

Then $\eta_{n,d} > 1$,

$$L_d = \log(en) + \log\eta_{n,d},$$

and $\tau_{n,d} < 1$ gives $\eta_{n,d} < L_d$. For every $\eta > 1$,

$$\log\eta = 2\log\sqrt{\eta} \leq 2(\sqrt{\eta} - 1) \leq \frac{\eta}{2},$$

where the last inequality is equivalent to $(\sqrt{\eta} - 2)^2 \geq 0$. Hence

$$\eta_{n,d} < \log(en) + \log\eta_{n,d} \leq \log(en) + \frac{\eta_{n,d}}{2},$$

so $\eta_{n,d} \leq 2\log(en)$, and therefore

$$L_d = \log(en) + \log\eta_{n,d} \leq \log(en) + \frac{\eta_{n,d}}{2} \leq 2\log(en).$$

Thus, throughout the case $\tau_{n,d} < 1$,

$$\frac{d}{n\log(en)} \leq 2\,\frac{d}{nL_d}.$$

Combining this with the square-root bound gives

$$\frac{\sqrt{d}}{n} + \frac{d}{n\log(en)} \leq 4\,\frac{d}{nL_d} = 4\min\left\{1, \tau_{n,d}\right\},$$

and hence the displayed comparison follows. Consequently,

$$\frac{c_\epsilon^{\mathrm{mf}}}{4}\min\left\{1, \frac{\sqrt{d}}{n} + \frac{d}{n\log(en)}\right\} \leq c_\epsilon^{\mathrm{mf}}\min\left\{1, \frac{d}{nL_d}\right\} \leq \mathfrak{R}_{n,d,\epsilon}.$$

For the upper bound, $d \geq 2$ gives $L_d \geq 1$, and therefore

$$\mathfrak{R}_{n,d,\epsilon} \leq C_\epsilon^{\mathrm{mf}}\min\left\{1, \frac{d}{nL_d}\right\} \leq C_\epsilon^{\mathrm{mf}}\frac{d}{n}.$$

Taking $c = c_\epsilon^{\mathrm{mf}}/4$ and $C = C_\epsilon^{\mathrm{mf}}$ proves the matched-frontier assertion in the proposition.

2. Fix n, d, ϵ, and an observed sample with $2 \le d < D_0$. Algorithm 1 selects its empirical-ratio branch whenever $d < D_0$. Hence, for that sample,

$$\widehat{V}_{n,d}^{\mathrm{JF}} = \widehat{V}_{n,d}^{\mathrm{ER}}.$$

3. Fix $\epsilon \in (0, 1/2)$ and a sequence (d_n) such that $d_n \ge 2$ for every n and $d_n \le d_{\max}$ for some finite $d_{\max}$. Then $d_{\max} \ge 2$. Define

$$c_{\epsilon,d_{\max}} = c_\epsilon^{\mathrm{mf}} \frac{2}{\log(ed_{\max})}, \qquad C_{\epsilon,d_{\max}} = C_\epsilon^{\mathrm{mf}}\, d_{\max}.$$

Since $\log(ed_{\max}) \ge 1$ and $d_{\max} \ge 2$,

$$\frac{2}{\log(ed_{\max})} \le d_{\max},$$

so

$$0 < c_{\epsilon,d_{\max}} \le C_{\epsilon,d_{\max}}.$$

For every $n \ge d_{\max}$, one has $d_n \le n$. With

$$L_{d_n} = \log(ed_n),$$

the bound $d_n \ge 2$ gives $L_{d_n} \ge 1$, and therefore

$$\frac{d_n}{nL_{d_n}} \le 1.$$

Thus the minimum in the matched frontier equals $d_n/(nL_{d_n})$. Monotonicity of the logarithm gives

$$L_{d_n} \le \log(ed_{\max}),$$

and, since $d_n \ge 2$,

$$\frac{2}{\log(ed_{\max})} \le \frac{d_n}{L_{d_n}}.$$

Therefore, for every $n \ge d_{\max}$,

$$\frac{c_{\epsilon,d_{\max}}}{n} \le c_\epsilon^{\mathrm{mf}} \frac{d_n}{nL_{d_n}} \le \mathfrak{R}_{n,d_n,\epsilon}.$$

Similarly,

$$\mathfrak{R}_{n,d_n,\epsilon} \le C_\epsilon^{\mathrm{mf}} \frac{d_n}{nL_{d_n}} \le C_\epsilon^{\mathrm{mf}} \frac{d_n}{n} \le \frac{C_{\epsilon,d_{\max}}}{n}.$$

These inequalities hold for every $n \ge d_{\max}$, hence for all sufficiently large n. $\square$

References

Susan Athey and Stefan Wager. Policy learning with observational data. *Econometrica*, 89(1): 133–161, 2021. doi: 10.3982/ECTA15732.

T. Tony Cai and Mark G. Low. Testing composite hypotheses, hermite polynomials and optimal estimation of a nonsmooth functional. *The Annals of Statistics*, 39(2):1012–1041, 2011. doi: 10.1214/10-AOS849.

Bibhas Chakraborty, Susan A. Murphy, and Victor J. Strecher. Inference for non-regular parameters in optimal dynamic treatment regimes. *Statistical Methods in Medical Research*, 19 (3):317–343, 2010. doi: 10.1177/0962280209105013.

Qizhao Chen, Morgane Austern, and Vasilis Syrgkanis. Inference on optimal dynamic policies via softmax approximation, 2023.

Ronald A. DeVore and George G. Lorentz. *Constructive Approximation*, volume 303 of *Grundlehren der mathematischen Wissenschaften*. Springer-Verlag, Berlin, 1993. ISBN 978-3-540-50627-0.

Yanjun Han, Jiantao Jiao, and Tsachy Weissman. Local moment matching: A unified methodology for symmetric functional estimation and distribution estimation under wasserstein distance. In *Proceedings of the 31st Conference on Learning Theory*, volume 75 of *Proceedings of Machine Learning Research*, pages 3189–3221. PMLR, 2018. URL https://proceedings.mlr.press/v75/han18b.html.

Keisuke Hirano and Jack R. Porter. Asymptotics for statistical treatment rules. *Econometrica*, 77(5):1683–1701, 2009. doi: 10.3982/ECTA6630.

Keisuke Hirano and Jack R. Porter. Impossibility results for nondifferentiable functionals. *Econometrica*, 80(4):1769–1790, 2012. doi: 10.3982/ECTA8681.

Jiantao Jiao, Kartik Venkat, Yanjun Han, and Tsachy Weissman. Minimax estimation of functionals of discrete distributions. *IEEE Transactions on Information Theory*, 61(5):2835–2885, 2015. doi: 10.1109/TIT.2015.2412945.

Jiantao Jiao, Yanjun Han, and Tsachy Weissman. Minimax estimation of the l_1 distance. *IEEE Transactions on Information Theory*, 64(10):6672–6706, 2018. doi: 10.1109/TIT.2018.2846245.

Toru Kitagawa and Aleksey Tetenov. Who should be treated? empirical welfare maximization methods for treatment choice. *Econometrica*, 86(2):591–616, 2018. doi: 10.3982/ECTA13288.

Lucien Le Cam. *Asymptotic Methods in Statistical Decision Theory*. Springer Series in Statistics. Springer-Verlag, New York, 1986. doi: 10.1007/978-1-4612-4946-7.

Alexander R. Luedtke and Mark J. van der Laan. Statistical inference for the mean outcome under a possibly non-unique optimal treatment strategy. *The Annals of Statistics*, 44(2): 713–742, 2016. doi: 10.1214/15-AOS1384.

Charles F. Manski. Statistical treatment rules for heterogeneous populations. *Econometrica*, 72 (4):1221–1246, 2004. doi: 10.1111/j.1468-0262.2004.00530.x.

Susan A. Murphy. Optimal dynamic treatment regimes. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 65(2):331–355, 2003. doi: 10.1111/1467-9868.00389.

Min Qian and Susan A. Murphy. Performance guarantees for individualized treatment rules. *The Annals of Statistics*, 39(2):1180–1210, 2011. doi: 10.1214/10-AOS864.

James M. Robins. A new approach to causal inference in mortality studies with a sustained exposure period—application to control of the healthy worker survivor effect. *Mathematical Modelling*, 7(9–12):1393–1512, 1986. doi: 10.1016/0270-0255(86)90088-6.

Paul R. Rosenbaum and Donald B. Rubin. The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70(1):41–55, 1983. doi: 10.1093/biomet/70.1.41.

Alexandre B. Tsybakov. *Introduction to Nonparametric Estimation*. Springer Series in Statistics. Springer, New York, 2009. ISBN 978-0-387-79052-7. doi: 10.1007/b13794.

Gregory Valiant and Paul Valiant. Estimating the unseen: Improved estimators for entropy and other properties. *Journal of the ACM*, 64(6):41:1–41:41, 2017. doi: 10.1145/3125643.

Mark J. van der Laan and Alexander R. Luedtke. Targeted learning of the mean outcome under an optimal dynamic treatment rule. *Journal of Causal Inference*, 3(1):61–95, 2015. doi: 10.1515/jci-2013-0022.

Haoyu Wei. Semiparametric off-policy inference for optimal policy values under possible non-uniqueness, 2025. Revised 2026.

Justin Whitehouse, Qizhao Chen, Morgane Austern, and Vasilis Syrgkanis. Inference on optimal policy values and other irregular functionals via softmax smoothing, 2025. Revised 2026.

Yihong Wu and Pengkun Yang. Minimax rates of entropy estimation on large alphabets via best polynomial approximation. *IEEE Transactions on Information Theory*, 62(6):3702–3720, 2016. doi: 10.1109/TIT.2016.2548468.

Zhenghao Zeng, Sivaraman Balakrishnan, Yanjun Han, and Edward H. Kennedy. Causal inference with high-dimensional discrete covariates, 2024. Revised 2026, arXiv version 3.