A minimax bracket for average treatment effect estimation with discrete adjustment and bounded heterogeneity

CausalSmith*

August 27, 2026

Abstract

This paper establishes finite-sample minimax bounds for a scalar average treatment effect in a real-outcome observed-data model with a finite discrete adjustment variable. The model imposes consistency, conditional exchangeability, and fixed overlap parameter ϵ; allows arbitrary cell masses; fixes the known outcome scale M, with conditional means in an interval of radius $M/2$; bounds conditional second central moments by M^2; and restricts the maximal cell-effect deviation to at most σM, where σ is the heterogeneity radius. For sample size n and alphabet size d, the number of adjustment cells, the paper constructs two clipped estimators and a clipped three-branch selector: a heavy–light signed Chebyshev factorial estimator for rare cells, an occupancy-weighted treated-control estimator for empirically crossed cells, and a constant-zero branch. The known-radius selector satisfies

$$\sup_{P \in \mathcal{P}_{d,\epsilon,M,\sigma}} E\{(\widehat{\tau}_n^{\star} - \tau(P))^2\} \leq C_\epsilon M^2 \left[\frac{1}{n} + \min\left\{1, \frac{d^2}{n^2 \log^2(en)}, \sigma^2 + \frac{d}{n^2}\right\}\right].$$

The same model class admits the lower benchmark

$$c_\epsilon M^2 \left[\frac{1}{n} + \min\left(1, \frac{d}{n^2}\right) + \sigma^2 \min\left\{1, \frac{d^2}{n^2 \log^2(en)}\right\}\right].$$

These two bounds match in order at exact homogeneity, at the unrestricted-radius endpoint, for every fixed positive radius, in the saturated alphabet regime, and at the parametric-dominance elbows. The results give a radius-indexed minimax bracket for point-estimation mean-squared error under fixed overlap, known outcome scale M, and known heterogeneity radius σ. The selector attaining the upper bound is a *known-radius* procedure: it uses σ as an input to choose deterministically between its polynomial, collision, and zero branches, and the regime theorem isolates the residual shrinking-radius region generated by the displayed upper and lower benchmarks.

1 Introduction

High-dimensional discrete adjustment creates a finite-sample tension that is absent from fixed-dimensional semiparametric calculations. The average treatment effect is identified by the usual

*Machine-generated by the CausalSmith research pipeline. The displayed formal statements and proofs are Lean-verified under the paper's theorem-local citation interface; the verification appendix records the exact scope, toolchain, and commit, including any statement that is conditional on a published input. Cited work otherwise supplies framing and positioning.

finite-cell adjustment formula under consistency, conditional exchangeability, and overlap, but estimation can be dominated by cells in which treated and control observations rarely meet. In binary-outcome discrete adjustment, Zeng et al. [2024] give collision-estimator bounds, exact-homogeneity upper and lower rates, and unrestricted binary-outcome lower bounds in their stated regimes.

This paper establishes a finite-sample minimax bracket for the same causal target with real outcomes under a radius-indexed finite-cell model. The adjustment variable takes d values with arbitrary cell masses; treatment is binary; outcomes are real-valued; overlap is fixed by ϵ; supported arm-cell conditional means lie in $[-M/2, M/2]$; supported arm-cell conditional second central moments are bounded by M^2; and the maximal supported-cell treatment-effect deviation from the average treatment effect is bounded by σM, where $\sigma \in [0, 2]$ is known to the estimator. The resulting model class is $\mathcal{P}_{d,\epsilon,M,\sigma}$ in Definition 1, and the risk criterion is the clipped-estimator minimax mean-squared risk in Definition 5. The bracket becomes an exact-order rate statement at the two radius endpoints, every fixed positive radius, saturation, and the parametric-dominance elbows; the regime theorem separately identifies the residual shrinking-radius region generated by the displayed bounds.

The constructive result combines two estimators that exploit different information sources. The polynomial estimator in Algorithm 1 uses a sample split, a pilot heavy–light classification, and a signed Chebyshev factorial statistic for light cells. Its risk bound has the nonparametric scale $d^2/\{n^2 \log^2(en)\}$, informed by the large-alphabet polynomial-estimation logic of Jiao et al. [2015], Wu and Yang [2016]. The collision estimator in Algorithm 2 averages treated-control contrasts over empirically crossed cells with occupancy weights; its risk bound contains the radius term σ^2 and the crossed-cell term d/n^2. The selector in Algorithm 3 chooses among these two estimators and the zero estimator using the known values of n, d, and σ.

The main upper bound, Theorem 2, shows that this selector satisfies

$$C_\epsilon M^2 \left[\frac{1}{n} + \min\left\{1, \frac{d^2}{n^2 \log^2(en)}, \sigma^2 + \frac{d}{n^2}\right\}\right]$$

up to the displayed overlap-dependent constant. The converse in Theorem 3 embeds binary hard experiments into the same real-outcome class. Its exact-homogeneity component supplies $M^2\{n^{-1} + \min(1, d/n^2)\}$, and its radius-channel component supplies

$$M^2\sigma^2 \min\left\{1, \frac{d^2}{n^2 \log^2(en)}\right\}.$$

Together these statements give the all-alphabet bracket in Theorem 4.

The all-parameter statement is a *bracket*: an upper and a lower benchmark for the same minimax risk over the same model class, valid at every admissible (n, d, M, σ). The bracket gives exact-order rates at exact homogeneity, the unrestricted-radius endpoint, every fixed positive radius, saturation, and the stated parametric-dominance elbows. Any divergence of the displayed upper-to-lower benchmark ratio is confined to the residual shrinking-radius wedge characterized in Theorem 5, which also exhibits a sequence inside that wedge.

The bracket has sharp implications across the named regimes. At $\sigma = 0$, Proposition 2 gives the exact-homogeneity rate $M^2\{n^{-1} + \min(1, d/n^2)\}$. At $\sigma = 2$, Lemma 1 identifies the radius-indexed class with the unrestricted-radius class, and the endpoint rate becomes $M^2\{n^{-1} + \min(1, d^2/[n^2 \log^2(en)])\}$. For every radius bounded away from zero, and in the saturation and parametric-dominance regimes, Theorem 5 shows that the minimax risk is equivalent to the selector benchmark. The same theorem isolates the residual shrinking-radius wedge where the displayed upper and lower benchmarks can separate.

The analysis is connected to three literatures. The causal setup follows the potential-outcomes and program-evaluation tradition of Rubin [1974, 1979], Rosenbaum and Rubin [1983], Hahn [1998], Robins et al. [1994], van der Laan and Robins [2003], Tsiatis [2006], Imbens and Wooldridge [2009], Imbens and Rubin [2015]. The high-dimensional treatment-effect literature, including Belloni et al. [2017], Chernozhukov et al. [2018], Wager and Athey [2018], Nie and Wager [2021], Semenova and Chernozhukov [2021], Kennedy [2022, 2023], Yadlowsky [2022], Celentano and Wainwright [2023], Jiang et al. [2025], motivates the role of nuisance complexity and heterogeneity in causal estimation. The polynomial branch draws on large-alphabet functional estimation and minimax testing tools from Paninski [2003], Valiant and Valiant [2011a,b, 2016, 2017], Jiao et al. [2015], Wu and Yang [2016, 2019], Le Cam [1986], Tsybakov [2009], Cai and Low [2011], Donoho et al. [1990], Timan [1963]. The verification appendix records the checked statement layer, the proof-audit status, and the attribution role of the citations.

The rest of the paper is organized as follows. The related-work section positions the selector benchmark against the published binary collision guarantee. The setup section defines the real-outcome model class, the average treatment effect, and the minimax risk. The estimator section introduces the polynomial estimator, the collision estimator, and the known-radius selector, then proves their upper bounds. The lower-bound section gives the affine and radius-channel embeddings from binary source experiments. The main-results section assembles the two-sided bracket and the matched-regime algebra. The discussion section records the residual shrinking-radius region and the scope of the bracket.

Theorem map. Four results carry the paper, and it may help to name them before the machinery arrives. Theorem 2 is the *upper bound*: the known-radius selector of Algorithm 3, built from the polynomial and collision estimators, has worst-case risk at most $C_\epsilon M^2 r_{n,d,\sigma}$ over the whole class. Here $r_{n,d,\sigma}$ is the selector benchmark for the constructed upper bound. Theorem 3 is the *lower bound*: the minimax risk is at least the capped converse benchmark

$$c_\epsilon M^2 \ell^{\mathrm{cap}}_{n,d,\sigma}, \qquad \ell^{\mathrm{cap}}_{n,d,\sigma} := n^{-1} + \min(1, d/n^2) + \sigma^2 \min\{1, d^2/[n^2 \log^2(en)]\},$$

proved by embedding binary hard experiments into the same real-outcome class through an affine map and a Bernoulli radius channel. Theorem 4 is the *bracket*, which puts the two together over a common index range. Theorem 5 is the *regime theorem*: it identifies the regimes where the selector and capped converse benchmarks agree in order, so that the bracket is a rate, and the residual shrinking-radius wedge where the displayed benchmarks separate. Proposition 2 specialises the bracket at $\sigma = 0$ and $\sigma = 2$. The remaining formal statements supply ingredients for these five statements or the comparison with the published binary results.

2 Related work

The paper belongs to the potential-outcomes tradition in which causal parameters are defined from potential responses and identified from observed data under consistency, conditional exchangeability, and overlap. Foundational work by Rubin [1974, 1979] and Rosenbaum and Rubin [1983] supplies the adjustment logic behind scalar average treatment effect analysis, while Imbens and Wooldridge [2009] and Imbens and Rubin [2015] develop the modern econometric formulation. In this language, the present analysis studies a real-outcome observational experiment with a finite discrete confounder and asks how accurately the average treatment effect can be estimated when the adjustment distribution may place highly uneven mass across cells.

Semiparametric work gives the classical benchmark for average treatment effect estimation when the nuisance functions can be estimated with enough regularity. The efficient influence-function calculations of Hahn [1998], the propensity-score weighting and series-regression estimators of Hirano et al. [2003], inverse-probability and augmented estimation ideas associated with Robins et al. [1994], and the treatments in van der Laan and Robins [2003], Tsiatis [2006], van der Vaart [1998] organize the fixed-dimensional theory around first-order regularity and variance. Three regimes should be kept apart. In fixed-dimensional efficient estimation the adjustment variable is held fixed while n grows, and the semiparametric efficiency bound is attained at rate $n^{-1/2}$, with the alphabet treated as fixed. In high-dimensional nuisance estimation, exemplified by Belloni et al. [2017], Chernozhukov et al. [2018], Kennedy [2022, 2023], parametric nuisance restrictions are replaced by rate and orthogonality conditions, and the alphabet enters through how well the nuisances can be learned. The regime studied here is a third one: the adjustment variable is exactly discrete, d grows with n, cell masses are unrestricted, and the binding constraint is combinatorial – whether a cell contains observations from both arms at all. That sparse-crossing constraint is why the risk here carries a d^2/n^2-type term. The present paper takes a complementary minimax perspective: the adjustment variable is exactly discrete, the number and masses of cells are part of the finite-sample problem, and the risk benchmark records the price of sparse treated-control overlap within cells.

A second line of work studies treatment-effect heterogeneity and its consequences for estimation. Methods for heterogeneous treatment effects and causal forests, such as Wager and Athey [2018], Nie and Wager [2021], and Semenova and Chernozhukov [2021], emphasize learning conditional effects or policy-relevant summaries. Related high-dimensional analyses study nuisance-estimation effects, debiasing phenomena in missing-data models, and AIPW variance behavior, including Yadlowsky [2022], Celentano and Wainwright [2023], and Jiang et al. [2025]. Here heterogeneity enters through a scalar radius that bounds the maximal cell-effect deviation around the target average effect. That radius indexes a family of same-class minimax problems: small radii favor estimators that borrow strength across cells, while larger radii make directly observed within-cell treatment-control contrasts more valuable.

The closest comparison is the discrete-adjustment minimax theory of Zeng et al. [2024]. Their analysis considers a binary-outcome discrete-confounder experiment and develops lower bounds and collision-based estimation benchmarks for unrestricted heterogeneity, exact homogeneity, and approximate homogeneity. Their collision phenomena identify the sparse-cell difficulty created by requiring both treatment arms to appear in the same adjustment cell, and their binary hard experiments provide the source ingredients for the embedding arguments used later in this paper. The present results transfer and extend that logic to real outcomes under bounded conditional means and conditional second central moments, arbitrary cell masses, fixed overlap, and a known maximal heterogeneity radius, yielding a radius-indexed bracket on one observational model class.

It is worth stating their rates explicitly, since they are the benchmark against which the present contribution should be read. In the binary-outcome discrete-adjustment model with fixed overlap, Zeng et al. [2024] show that the usual regression, weighting, and doubly robust estimators have mean-squared error of order

$$\frac{d^2}{n^2} + \frac{1}{n},$$

whereas the minimax lower benchmark for unrestricted heterogeneity is the strictly smaller

$$\frac{d^2}{n^2 \log^2 n} + \frac{1}{n},$$

so the minimax benchmark improves on the standard-estimator rate in the sparse-cell regime. Under exact effect homogeneity the rate improves to

$$\frac{d}{n^2} + \frac{1}{n},$$

and their occupancy-weighted collision estimator, under approximate homogeneity at binary radius σ_{bin}, carries the algebraic remainder

$$\sigma_{\text{bin}}^2 + \frac{d}{n^2}$$

together with an exponentially small adverse-event term.

The comparison is easiest to read side by side. The two analyses share a causal target and an overlap condition and differ in outcome type, in how heterogeneity is indexed, and in what is proved.

	Zeng et al. [2024]	This paper
Outcome	binary	real, second moment $\leq M^2$
Cell masses	arbitrary	arbitrary
Overlap	fixed ϵ	fixed ϵ
Heterogeneity	three separate regimes	one radius $\sigma \in [0, 2]$
Estimator	collision	selector over polynomial, collision, zero
Upper	$\sigma_{\text{bin}}^2 + d/n^2$	selector benchmark $M^2 r_{n,d,\sigma}$
Lower, unrestricted	$d^2/(n^2 \log^2 n) + 1/n$	$\sigma = 2$ endpoint $M^2\{n^{-1} + \min(1, d^2/[n^2 \log^2($
Lower, exact homogeneous	$d/n^2 + 1/n$	$\sigma = 0$ endpoint $M^2\{n^{-1} + \min(1, d/n^2)\}$
Lower, radius indexed	benchmark suite by regime	$M^2\{n^{-1} + \min(1, d/n^2) + \sigma^2 \min(1, d^2/[n^2 \log$
Radius-indexed bracket	binary benchmark suite by regime	one real-outcome class, same indices

Three entries carry the novelty. The outcome row is why the estimators had to be rebuilt: a second-moment envelope supplies variance control, so every statistic here is clipped and every risk bound runs through variances. The heterogeneity row is why a bracket is meaningful at all: with σ indexing one nested family, the binary exact-homogeneity and unrestricted rates become the $\sigma = 0$ and $\sigma = 2$ endpoints of a single statement. The split lower rows display the unrestricted, exact-homogeneity, and radius-indexed lower benchmarks separately. The final row records the same-class feature of the present bracket: both benchmarks are stated over one real-outcome class indexed by the same n, d, M, σ. Their upper-to-lower ratio therefore measures the remaining benchmark gap for one minimax problem, and it becomes an exact-order rate statement precisely in the regimes where Theorem 5 proves the benchmarks comparable.

The present use of Zeng et al. [2024] has three distinct layers. Their work supplies the binary hard-experiment mechanisms for unrestricted and exact-homogeneity converses, the sparse crossed-cell collision mechanism, and the binary collision benchmark used for comparison. In the notation of this paper, the corresponding binary source statements are recorded as Definition 7 and Lemmas 3 to 6. The new real-outcome analysis constructs the heavy–light polynomial estimator, the continuous-outcome occupancy estimator, the known-radius selector, the affine and radius-channel transfers into $\mathcal{P}_{d,\epsilon,M,\sigma}$, and the all-alphabet same-class bracket. In that bracket, the binary exact-homogeneity and unrestricted scales appear as the $\sigma = 0$ and $\sigma = 2$ endpoints of one radius-indexed family, and Proposition 3 records the algebraic comparison with the binary collision remainder $\sigma^2_{\text{bin}} + d/n^2$.

The construction also connects to large-alphabet functional estimation. Polynomial approximation and unbiased or nearly unbiased factorial statistics have been central in estimating distributional functionals with many rare categories, as in Paninski [2003], Valiant and Valiant [2011a,b, 2016, 2017], Jiao et al. [2015], Wu and Yang [2016], and Wu and Yang [2019]. The estimator developed below uses that same rare-category logic for a causal target: light adjustment cells are handled through a polynomial device, while heavier empirically crossed cells are handled by an occupancy-weighted treated-control contrast. Classical tools for minimax lower bounds and approximation, including Le Cam [1986], Tsybakov [2009], Cai and Low [2011], Donoho et al. [1990], Timan [1963], and moment inequalities for symmetric statistics, including Hoeffding [1948], Serfling [1980], provide the surrounding mathematical vocabulary.

The empirical setting that motivates discrete adjustment at this scale is the 401(k) eligibility literature, where participation and financial-wealth outcomes are adjusted for finely stratified income and demographic cells [Poterba et al., 1994, 1995]. That application is the consumer of the binary analysis of Zeng et al. [2024], and it motivates the real-outcome risk bounds pursued here: wealth is naturally real-valued, and the cell counts in such designs are exactly the sparse regime in which the crossed-cell term matters.

Finally, the paper is related to higher-order and robust estimation ideas in semiparametric causal inference. Higher-order influence-function methods of Robins et al. [2008, 2017] and robust mean-estimation techniques such as Lugosi and Mendelson [2019] illustrate how weak moment information and complex nuisance structure can change attainable risk. The contribution here is a finite-alphabet minimax bracket for scalar average treatment effect estimation under fixed overlap and a bounded heterogeneity radius. The upper side is attained by a known-radius selector over polynomial, collision, and trivial estimators; the lower side is transported from binary experiments into the same real-outcome model. The resulting benchmarks match at the exact-homogeneity and unrestricted-radius endpoints, at every fixed positive radius, in the saturated alphabet regime, and at the parametric-dominance elbows.

3 Setup and assumptions

Throughout, n denotes sample size and d denotes the number of adjustment cells. Cell indices such as k range over $\{1, \ldots, d\}$, and treatment levels such as a range over $\{0, 1\}$. Constants c_ϵ and C_ϵ may change from line to line and depend only on the overlap parameter ϵ. The outcome scale is $M \geq 1$, and the heterogeneity radius is $\sigma \in [0, 2]$. We use $O(\cdot)$, $\lesssim_\epsilon$, and $\asymp_\epsilon$ with constants depending only on ϵ.

A full-data law P governs the finite-cell observational unit with confounder X, binary treatment A, potential outcomes $Y(a)$, and observed real outcome Y. For a supported cell, p_k is the cell mass, π_k is the propensity, and μ_{ak} is the arm-cell conditional mean. The corresponding cell treatment effect is τ_k, the target average treatment effect is $\tau(P)$, and the centered cell-effect deviation is δ_k. These objects are fixed formally below before the minimax experiment is introduced.

The analysis starts from the potential-outcomes conditions for finite-cell adjustment, followed by scale and heterogeneity restrictions that make the finite-sample risk problem comparable across real-outcome laws.

Assumption 1. [`ass:consistency`] (Consistency) *Under P,*

$$Y = Y(A)$$

almost surely.

Consistency links the observed outcome to the potential outcome under the realized treatment assignment, following the potential-outcomes setup used in discrete-adjustment analyses such as Zeng et al. [2024].

Assumption 2. [`ass:conditional-exchangeability`] (Conditional exchangeability) *Under P,*

$$(Y(0), Y(1)) \perp A \mid X.$$

Conditional exchangeability is the selection-on-observables condition after conditioning on the discrete adjustment cell [Zeng et al., 2024]. Together with consistency, it gives the usual causal interpretation of cell-level treated-control contrasts [Rubin, 1974, Rosenbaum and Rubin, 1983, Imbens and Rubin, 2015].

Assumption 3. [`ass:overlap`] (Fixed overlap) *For each cell k, let*

$$p_k := P(X = k),$$

and, when $p_k > 0$, let

$$\pi_k := P(A = 1 \mid X = k).$$

For every k with $p_k > 0$,

$$\epsilon \leq \pi_k \leq 1 - \epsilon.$$

The fixed-overlap requirement imposes strong overlap on every supported cell, as in binary discrete-adjustment work such as Zeng et al. [2024]. It keeps both treatment arms available in population within each cell while allowing the cell masses themselves to be arbitrary.

Assumption 4. [`ass:mean-normalization`] (Mean normalization) *The law P carries an arm-cell outcome distribution for every arm $a \in \{0, 1\}$ and* every *cell k, zero-mass cells included; let μ_{ak} denote its mean. Whenever the arm-cell probability $P(A = a, X = k)$ is positive, μ_{ak} is the observed conditional mean $E_P[Y \mid A = a, X = k]$. Only supported cells are constrained: for every $a \in \{0, 1\}$ and every k with $p_k > 0$,*

$$|\mu_{ak}| \leq M/2.$$

Mean normalization is the affine-centered bounded conditional-mean analogue of the bounded binary response scale in Zeng et al. [2024]. The centering fixes the outcome scale for the minimax comparison, and the restriction applies on the supported part of the adjustment distribution.

Assumption 5. [ass:approximate-homogeneity] (Approximate homogeneity) *For* every *cell k, let*

$$\tau_k := \mu_{1k} - \mu_{0k}$$

and

$$\delta_k := \tau_k - \sum_j p_j \tau_j,$$

both well defined on zero-mass cells as well. The cell-effect deviations satisfy

$$\max_{k: p_k > 0} |\delta_k| \le \sigma M.$$

Approximate homogeneity is the outcome-scale-normalized analogue of a maximal cell-effect-deviation radius in binary discrete adjustment [Zeng et al., 2024]. The radius σM measures the largest supported-cell departure of a cell treatment effect from the population-weighted average effect.

Assumption 6. [ass:second-central-moment] (Second central moment) *For every $a \in \{0, 1\}$ and every k with $p_k > 0$,*

$$E_P\left[(Y - \mu_{ak})^2 \mid A = a, X = k\right] \le M^2.$$

The moment bound extends the binary bounded-outcome scale to real-valued outcomes through a uniform conditional second-central-moment envelope. It keeps the noise level on the same scale as the conditional-mean and heterogeneity restrictions.

Role of the assumptions. It is worth pausing on what each condition is doing, because the combination is unusual. Fixed overlap ϵ is the standard identification requirement, but here it plays a second, finite-sample role: it guarantees that a cell which is occupied at all has a non-vanishing chance of containing both arms, which is exactly the event the collision estimator lives on. The bound $|\mu_{ak}| \le M/2$ is a scale normalization: it fixes the scale against which a treatment effect is judged large, and a finite minimax comparison needs such a scale because multiplying all outcomes by a constant multiplies the risk by its square. The conditional second-moment envelope $E[(Y - \mu_{ak})^2 \mid A = a, X = k] \le M^2$ is where this model departs from the binary-outcome literature. Bounded outcomes supply bounded or sub-Gaussian concentration; a second-moment envelope supplies variance control, so the estimators are clipped and the risk analysis runs through variances. That is the price of allowing real outcomes, and it is why both estimators here are total, clipped maps. Finally, the radius σ is the object of interest: it interpolates between exact effect homogeneity, where the population contrast can be read off any single cell, and unrestricted heterogeneity, where every cell must be estimated separately. Making σ an index of one nested family allows a single bracket to describe all of these regimes.

These conditions define two law classes. The first imposes the radius bound, while the second records the same overlap and moment envelope with the radius left free.

Definition 1. [def:model-class] (Real-outcome model class $\mathcal{P}_{d,\epsilon,M,\sigma}$)

$$\mathcal{P}_{d,\epsilon,M,\sigma} := \left\{ P : \begin{array}{l} X \in \{1, \dots, d\},\ A \in \{0, 1\},\ Y, Y(0), Y(1) \in \mathbb{R}, \\ 0 < \epsilon < 1/2, \quad M \ge 1, \quad 0 \le \sigma \le 2, \\ P \textit{ satisfies Assumptions } 1 \textit{to } 6 \end{array} \right\}.$$

The radius-indexed class $\mathcal{P}_{d,\epsilon,M,\sigma}$ is the main parameter space of the paper: all finite-cell masses are allowed, and the scalar radius records the amount of cell-effect variation available to the estimator.

Definition 2. [def:unrestricted-class] (Unrestricted-radius class $\mathcal{P}^{\mathrm{unr}}_{d,\epsilon,M}$)

$$\mathcal{P}^{\mathrm{unr}}_{d,\epsilon,M} := \left\{ P : \begin{array}{l} X \in \{1,\dots,d\},\ A \in \{0,1\},\ Y, Y(0), Y(1) \in \mathbb{R}, \\ 0 < \epsilon < 1/2, \quad M \geq 1, \\ P \textit{ satisfies Assumptions } 1 \textit{to } 4 \textit{and } 6 \end{array} \right\}.$$

The unrestricted-radius class $\mathcal{P}^{\mathrm{unr}}_{d,\epsilon,M}$ supplies the endpoint comparison for the radius scale under the same fixed-overlap and real-outcome moment envelope.

The observed-data experiment is generated by independent sampling from the observed margin. Write the observed record as O and its observed-data law as P_O.

Definition 3. [def:sample-experiment] (Sample experiment $\mathcal{E}_{n,d,\epsilon,M,\sigma}$) *Let*

$$O := (X, A, Y) \in \mathcal{O} := \{1,\dots,d\} \times \{0,1\} \times \mathbb{R},$$

and let P_O denote the observed-data law of O under P. The n-sample observed-data experiments generated by $\mathcal{P}_{d,\epsilon,M,\sigma}$ are

$$\mathcal{E}_{n,d,\epsilon,M,\sigma} := \left\{ Q_P^{(n)} = P_O^{\otimes n} : P \in \mathcal{P}_{d,\epsilon,M,\sigma} \right\}.$$

Thus each sampling law $Q_P^{(n)}$ belongs to the experiment $\mathcal{E}_{n,d,\epsilon,M,\sigma}$ induced by a law in $\mathcal{P}_{d,\epsilon,M,\sigma}$.

The target parameter is the finite-cell average treatment effect. Its definition uses the same cell masses and arm-cell conditional means introduced in the assumptions.

Definition 4. [def:ate-functional] (Average treatment effect $\tau(P)$) *For a real-outcome finite-cell law P indexed by $k = 1,\dots,d$ and an overlap parameter ϵ with $0 < \epsilon < 1/2$, assume consistency, conditional exchangeability of $(Y(0), Y(1))$ and A given X for measurable potential-outcome events, overlap $\epsilon \leq \pi_k \leq 1-\epsilon$ whenever $p_k > 0$, and first-moment integrability of each arm-cell outcome law for every $a \in \{0,1\}$ and positive-mass cell. With p_k the cell mass and μ_{ak} the arm-cell conditional outcome mean, define*

$$\tau(P) := \sum_{k=1}^{d} p_k(\mu_{1k} - \mu_{0k}).$$

Under these conditions $\tau(P)$ coincides with the full-data causal contrast $\int \{Y(1) - Y(0)\}\, dP$ and is identified from the observed margin by the standard finite-cell g-computation argument [Rubin, 1974, Rosenbaum and Rubin, 1983, Robins et al., 1994, Imbens and Rubin, 2015]; the minimax analysis targets this observed-data functional throughout.

For the risk statements, estimators are clipped to the natural scale implied by the mean normalization. The selector benchmark used later is

$$r_{n,d,\sigma} := n^{-1} + \min\left\{ \sigma^2 + \frac{d}{n^2}, \frac{d^2}{n^2 \log^2(en)}, 1 \right\},$$

where $r_{n,d,\sigma}$ records the parametric term together with the best of the collision, polynomial, and zero-estimator scales.

Definition 5. [`def:minimax-risk`] (Minimax risk $\mathsf{R}_{n,d,\epsilon,M,\sigma}$ and estimator class $\mathcal{T}_{n,M}$) *Let*

$$\mathcal{T}_{n,M} := \big\{T : \mathcal{O}^n \to [-M, M] \ : \ T \text{ is measurable}\big\}.$$

The minimax mean-squared risk over $\mathcal{P}_{d,\epsilon,M,\sigma}$ is

$$\mathsf{R}_{n,d,\epsilon,M,\sigma} := \inf_{T\in\mathcal{T}_{n,M}} \sup_{P\in\mathcal{P}_{d,\epsilon,M,\sigma}} E_{Q_P^{(n)}}\left[(T - \tau(P))^2\right].$$

The estimator class $\mathcal{T}_{n,M}$ matches the range of the target under the scale bound, and the minimax risk $\mathsf{R}_{n,d,\epsilon,M,\sigma}$ is the paper's central finite-sample loss criterion.

The main rate displays use the following benchmark notation. The table gives each symbol one role; later sections keep these names fixed.

Symbol	Name	Definition
$u_{n,d}$	rare-cell polynomial remainder	$d^2/\{n^2\log^2(en)\}$
$h_{n,d,\sigma}$	collision remainder	$\sigma^2 + d/n^2$
$b_{n,d}$	exact-homogeneity baseline	$n^{-1} + d/n^2$
$r_{n,d,\sigma}$	selector benchmark	$n^{-1} + \min\{1, u_{n,d}, h_{n,d,\sigma}\}$
$q_{n,d,\sigma}$	triangular selector benchmark	$n^{-1} + \min\{u_{n,d}, \sigma^2 + d/n^2\}$
$\ell^{\mathrm{cap}}_{n,d,\sigma}$	capped converse benchmark	$n^{-1} + \min(1, d/n^2) + \sigma^2\min(1, u_{n,d})$
$\ell^{\mathrm{tri}}_{n,d,\sigma}$	triangular product benchmark	$b_{n,d} + \sigma^2 u_{n,d}$

The word "frontier" in the verified labels names the selector benchmark $r_{n,d,\sigma}$. It records the upper benchmark achieved by the constructed known-radius selector in Theorem 2. The capped converse benchmark $\ell^{\mathrm{cap}}_{n,d,\sigma}$ is the lower benchmark in the all-alphabet bracket, while $\ell^{\mathrm{tri}}_{n,d,\sigma}$ is the product-form shorthand used in the triangular benchmark algebra. Some anchored theorem statements introduce local symbols inside their displays; the reader-facing prose uses the superscripted notation above whenever both lower benchmarks appear in the same discussion. Wherever $r_{n,d,\sigma}$ and $\ell^{\mathrm{cap}}_{n,d,\sigma}$ agree in order, the selector benchmark is the minimax rate; the residual shrinking-radius region is the localized regime where the displayed selector and product-form converse benchmarks separate.

The normalization also pins down the unrestricted endpoint of the radius scale.

Lemma 1. [`lem:scale-sanity`] (Scale sanity) *For any $d \in \mathbb{N}$ and any $\epsilon, M \in \mathbb{R}$, let $\mathcal{P}^{\mathrm{unr}}_{d,\epsilon,M}$ and $\mathcal{P}_{d,\epsilon,M,\sigma}$ be the model classes in Definitions 1 and 2. For each law P, write*

$$\tau_k(P) = \mu_{1k}(P) - \mu_{0k}(P), \qquad \tau(P) = \sum_{k\in\{1,\dots,d\}} p_k(P)\tau_k(P), \qquad \delta_k(P) = \tau_k(P) - \tau(P).$$

Every $P \in \mathcal{P}^{\mathrm{unr}}_{d,\epsilon,M}$ satisfies

$$|\tau_k(P)| \le M \quad \text{for every } k \text{ with } p_k(P) > 0, \qquad |\tau(P)| \le M,$$

and

$$|\delta_k(P)| \le 2M \quad \text{for every } k \text{ with } p_k(P) > 0.$$

Consequently the two classes have the same underlying laws at radius 2: for every $P \in \mathcal{P}^{\mathrm{unr}}_{d,\epsilon,M}$ there exists $Q \in \mathcal{P}_{d,\epsilon,M,2}$ with the same law as P, and for every $Q \in \mathcal{P}_{d,\epsilon,M,2}$ there exists $P \in \mathcal{P}^{\mathrm{unr}}_{d,\epsilon,M}$ with the same law as Q.

The proof is deferred to Section E.

Lemma 1 makes the radius normalization exact: the bounded conditional means imply $|\tau(P)| \le M$, supported cell effects lie within the same outcome scale, and radius 2 coincides with the unrestricted-radius law class. This calibration is used when the main bracket is specialized to endpoint regimes.

Finally, the lower-bound arguments later use binary source experiments from Zeng et al. [2024]. We record the source target and source law classes here so that the subsequent affine embeddings have a fixed reference point.

Definition 6. [`synth_1`] (Binary exact-homogeneity source target) *For a binary-outcome fixed-sample law P in the uniform-mass exact-homogeneity source family, let μ_{ak} denote the conditional response mean in arm $a \in \{0,1\}$ and cell $k \in \{1,\dots,d\}$. The source target is*

$$\psi(P) := \frac{1}{d}\sum_{k=1}^{d}(\mu_{1k} - \mu_{0k}).$$

Definition 7. [`synth_2`] (Binary source law classes) *For an overlap parameter $0 < \epsilon < 1/2$, let $\mathcal{D}_d^{\mathrm{bin}}(\epsilon)$ be the class of binary d-cell full-data laws for $(X, A, B(0), B(1), B)$, with $X \in \{1,\dots,d\}$, $A, B(0), B(1), B \in \{0,1\}$, consistency $B = B(A)$, conditional exchangeability $(B(0), B(1)) \perp A \mid X$, arbitrary cell masses p_k, propensities π_k satisfying $\epsilon \le \pi_k \le 1-\epsilon$ on every positive-mass cell, and arm means $\mu_{ak}^{\mathrm{bin}} = E[B \mid A = a, X = k] \in [0,1]$. Define its uniform-mass exactly homogeneous subclass by*

$$\mathcal{H}_d^{\mathrm{bin}}(\epsilon) := \left\{ P \in \mathcal{D}_d^{\mathrm{bin}}(\epsilon) : p_k(P) = \frac{1}{d} \text{ for every } k,\ \mu_{1k}^{\mathrm{bin}}(P) - \mu_{0k}^{\mathrm{bin}}(P) = \mu_{1\ell}^{\mathrm{bin}}(P) - \mu_{0\ell}^{\mathrm{bin}}(P) \text{ for all } k, \ell \right\}.$$

The source target in Definition 6 and the binary classes in Definition 7 provide the binary ingredients transported in the lower-bound section. Their role is to specify the published discrete-adjustment experiments before the paper embeds them into the real-outcome class $\mathcal{P}_{d,\epsilon,M,\sigma}$.

4 Estimators and upper bounds

We now turn from identification and the minimax loss criterion to constructive estimation. The upper bound is built from two complementary procedures. The first estimator, T_n^{poly}, uses a sample split to separate cell classification from estimation and applies a Chebyshev polynomial device to rare cells, following the large-alphabet functional-estimation logic of Jiao et al. [2015], Wu and Yang [2016]. The second estimator, T_n^{col}, forms treatment-control contrasts in empirically crossed cells and weights those contrasts by their observed occupancy. The final estimator, $\widehat{\tau}_n^{\star}$, uses the known radius to choose among the polynomial estimator, the collision estimator, and the zero estimator.

The polynomial construction is useful when many cells are light. Its pilot block identifies cells with enough empirical mass for a plug-in treated-control contrast, while its estimation block evaluates both the heavy-cell plug-in terms and the light-cell factorial terms. The ordered distinct-index structure is the finite-sample analogue of unbiased polynomial estimation for sparse categorical functionals, with the sample split keeping the random heavy-light classification separate from the contribution estimates [Hoeffding, 1948, Serfling, 1980].

The proof of the upper bound follows this construction. Conditional on the pilot block, the heavy set is deterministic. Heavy cells are controlled by stratified empirical means and fixed overlap; light cells are controlled by shifted-Chebyshev coefficient envelopes and covariance bounds for one-mark factorial statistics. Summing the conditional risk and then averaging over the pilot block gives the polynomial risk bound. The collision branch is analyzed separately by decomposing its error into ordinary arm-cell noise and the bias contributed by uncrossed cells; the approximate-homogeneity radius controls the latter term, and the occupancy reciprocal bound controls the random denominator.

Algorithm 1 (Polynomial estimator T_n^{poly})

For observations $O_i = (X_i, A_i, Y_i)$, with $X_i \in \{1, \dots, d\}$, $A_i \in \{0, 1\}$, and $Y_i \in \mathbb{R}$, fix a polynomial handle (N, ρ) with $\rho > 0$. Let $L := \log(en)$,

$$D_6 := 8 \log(27/4), \qquad \alpha_0 := \min\left\{1, \frac{1}{64 \log 6}, \frac{1}{512 D_6}\right\}, \qquad K_n := \lfloor \alpha_0 L \rfloor .$$

The handle certifies that $2 \le K_n$ whenever $0 < n$, $0 < d$, $N \le n$, and $d \le \rho n L$. The active branch below is used on $N \le n$ and $d \le \rho n L$; the calibrated value is $T_n^{\mathrm{poly}} = 0$ when $N > n$ or $d > \rho n L$.

1. Split the sample into the pilot and estimation blocks

$$I_0 := \{1, \dots, \lfloor n/2 \rfloor\}, \qquad I_1 := \{\lfloor n/2 \rfloor + 1, \dots, n\}, \qquad m_1 := |I_1| = n - \lfloor n/2 \rfloor.$$

2. Use I_0 to form the empirically heavy and light cell sets

$$\widehat{\mathcal{H}} := \left\{ k : 256L < \sum_{i \in I_0} \mathbf{1}\{X_i = k\} \right\}, \qquad \widehat{\mathcal{L}} := \{1, \dots, d\} \setminus \widehat{\mathcal{H}}.$$

3. On each heavy cell $k \in \widehat{\mathcal{H}}$, use I_1 to compute the estimation-block arm counts, cell count, arm outcome totals, totalized arm means, and normalized plug-in contribution

$$N_{ak}^{(1)} := \sum_{i \in I_1} \mathbf{1}\{X_i = k, A_i = a\}, \qquad N_k^{(1)} := \sum_{i \in I_1} \mathbf{1}\{X_i = k\},$$

$$S_{ak}^{(1)} := \sum_{i \in I_1} \mathbf{1}\{X_i = k, A_i = a\} Y_i, \qquad \overline{Y}_{ak}^{(1)} := \begin{cases} S_{ak}^{(1)} / N_{ak}^{(1)}, & N_{ak}^{(1)} > 0, \\ 0, & N_{ak}^{(1)} = 0, \end{cases}$$

$$\widehat{h}_k := \frac{N_k^{(1)}}{m_1} \frac{\overline{Y}_{1k}^{(1)} - \overline{Y}_{0k}^{(1)}}{M}.$$

4. On each light cell $k \in \widehat{\mathcal{L}}$, set

$$B := \frac{4096L}{m_1}, \qquad g_{K_n,j} := (-1)^j \frac{2^{2j+3}}{K_n(K_n + j + 2)} \binom{K_n + j + 2}{2j + 4}, \qquad j = 0, \dots, K_n - 2.$$

With $(m)_r$ denoting the falling factorial, evaluate the ordered distinct-index one-mark statistic

$$U_{k,a,j} := \frac{1}{(m_1)_{j+2}} \sum_{\substack{(i_0,\ldots,i_{j+1})\in I_1^{j+2} \\ i_0,\ldots,i_{j+1} \text{ distinct}}} \frac{Y_{i_0}}{M} \mathbf{1}\{X_{i_0} = k, A_{i_0} = a\} \left(\prod_{\ell=1}^{j} \mathbf{1}\{X_{i_\ell} = k, A_{i_\ell} = a\} \right) \mathbf{1}\{X_{i_{j+1}} = k\},$$

and set the normalized light-cell contribution to

$$\widehat{P}_k := \sum_{a\in\{0,1\}} (-1)^{1-a} \sum_{j=0}^{K_n-2} \frac{g_{K_n,j}}{B^{j+1}} U_{k,a,j}.$$

5. Writing $\mathrm{clip}_{[a,b]}(x) := \min\{b, \max\{a, x\}\}$ for truncation to the interval $[a, b]$, sum the heavy and light contributions and output the clipped active-branch statistic

$$T_n^{\mathrm{poly}} := M\, \mathrm{clip}_{[-1,1]} \left(\sum_{k\in\widehat{\mathcal{H}}} \widehat{h}_k + \sum_{k\in\widehat{\mathcal{L}}} \widehat{P}_k \right).$$

The factorial display in Algorithm 1 is a mathematical representation of the statistic. A direct implementation first scans the estimation block once to store, for each cell and arm, the counts $N_{ak}^{(1)}$, totals $S_{ak}^{(1)}$, and powers or falling-factorial count products needed for orders $0, \ldots, K_n - 2$. The light-cell terms are then evaluated from these aggregated arrays, so the program sums over cells, arms, and polynomial orders rather than over ordered tuples of observations. With this aggregation, Lemma 13 gives the post-aggregation operation count $C_{\mathrm{op}}(d + 1)(K_n + 1)^2$, with C_{op} independent of n, d, and M. Including the raw scans over the pilot and estimation blocks, the computational cost is $O\{n + d(K_n + 1)^2\}$ up to constants depending on the fixed overlap calibration.

Concretely, the whole procedure is five passes over arrays, and it is worth spelling them out because the ordered-tuple display above looks more expensive than the computation actually is.

1. *Pilot pass.* Scan I_0 once and accumulate $c_k = \#\{i \in I_0 : X_i = k\}$ for each cell. Mark cell k heavy if $c_k > 256L$ and light otherwise. This fixes $\widehat{\mathcal{H}}$ and $\widehat{\mathcal{L}}$ before any outcome on I_1 is touched, which is what keeps the two blocks independent.

2. *Aggregation pass.* Scan I_1 once and accumulate, for each cell k and arm a, the count $N_{ak}^{(1)}$ and the outcome total $S_{ak}^{(1)}$; also accumulate the cell count $N_k^{(1)}$. The aggregate arrays carry the sample information used by the later arithmetic steps, so the ordered tuples in the display are represented through counts and totals.

3. *Heavy cells.* For $k \in \widehat{\mathcal{H}}$, form $\overline{Y}_{ak}^{(1)} = S_{ak}^{(1)}/N_{ak}^{(1)}$, taking the value 0 when $N_{ak}^{(1)} = 0$, and set $\widehat{h}_k = (N_k^{(1)}/m_1)(\overline{Y}_{1k}^{(1)} - \overline{Y}_{0k}^{(1)})/M$. The zero-count convention is what makes the statistic total rather than partially defined.

4. *Light cells.* For $k \in \widehat{\mathcal{L}}$, the ordered-tuple sum defining $U_{k,a,j}$ collapses to a closed form in the aggregates already stored. Summing over ordered distinct indices with a marked

leading index, a run of j further indices in the same arm-cell, and one trailing index in the cell gives

$$U_{k,a,j} = \frac{S_{ak}^{(1)}}{M} \cdot \frac{\left(N_{ak}^{(1)} - 1\right)_j \left(N_k^{(1)} - j - 1\right)}{(m_1)_{j+2}},$$

where $(N)_r = N(N-1)\cdots(N-r+1)$ is the falling factorial and any factor with a negative argument is read as zero. The aggregate form represents the tuple sum directly: the leading index contributes the arm-cell outcome total $S_{ak}^{(1)}$, the j middle indices contribute a falling factorial of the remaining arm-cell count, and the trailing index contributes the remaining cell count. Build the falling factorials incrementally, $(N)_{r+1} = (N)_r\,(N-r)$, so each order $j = 0, \ldots, K_n - 2$ costs one multiplication; then accumulate $\widehat{P}_k = \sum_a (-1)^{1-a} \sum_j (g_{K_n,j}/B^{j+1}) U_{k,a,j}$. This is the $(d+1)(K_n+1)^2$ term in the operation count.

5. *Output.* Sum $\widehat{h}_k$ over heavy cells and $\widehat{P}_k$ over light cells, clip the total to $[-1, 1]$, and multiply by M. The clipping gives the estimator its $[-M, M]$ range and hence the worst-case risk control used when the polynomial branch is evaluated outside its calibrated range.

Outside the calibrated range $N_\epsilon \le n$ and $d \le \rho_\epsilon n \log(en)$ the procedure returns $T_n^{\mathrm{poly}} = 0$, so the estimator is defined across the whole sample space. The theorem fixes an overlap-dependent calibration handle $(N_\epsilon, \rho_\epsilon)$, and the selector uses the known indices n, d, M, σ: M sets the clipping level, σ determines the branch comparison in Algorithm 3, and n, d determine the polynomial degree and the displayed remainders. The heavy-light threshold and calibration handle are fixed by ϵ; the branch choice is deterministic in the model indices and the observed sample enters only through the two candidate statistics.

The collision estimator targets the part of the sample in which treated and control observations meet within the same adjustment cell. Its form is close to the occupancy statistics used in sparse-cell average treatment effect analysis and higher-order causal estimation: the sample supplies both the within-cell contrast and the empirical weight assigned to crossed cells [Robins et al., 2008, 2017]. Robust mean-estimation work under weak moment conditions, such as Lugosi and Mendelson [2019], provides a separate point of comparison for how moment assumptions shape attainable risk.

Algorithm 2 (Collision estimator T_n^{col})

For observations $O_1, \ldots, O_n$, define T_n^{col} by the following one-pass cell aggregation followed by one pass through the occupied categories.

1. For each arm $a \in \{0, 1\}$ and cell $k \in \{1, \ldots, d\}$, define the full-sample arm-cell count and outcome total

$$N_{ak} := \sum_{i=1}^{n} \mathbf{1}\{X_i = k, A_i = a\}, \qquad S_{ak} := \sum_{i=1}^{n} \mathbf{1}\{X_i = k, A_i = a\} Y_i.$$

2. Define the full-sample cell count, crossed-cell indicator, and total count in empirically crossed cells

$$N_k := N_{0k} + N_{1k}, \qquad I_k := \mathbf{1}\{N_{0k} N_{1k} > 0\}, \qquad R_n := \sum_{k=1}^{d} N_k I_k.$$

3. For positive arm-cell counts, define the arm-cell sample means

$$\overline{Y}_{ak} := \frac{S_{ak}}{N_{ak}} \qquad \text{when } N_{ak} > 0.$$

4. Output the occupancy-weighted treated-control statistic

$$T_n^{\mathrm{col}} := \mathrm{clip}_{[-M,M]}\left(\frac{\mathbf{1}\{R_n > 0\}}{R_n}\sum_{k=1}^{d} N_k I_k(\overline{Y}_{1k} - \overline{Y}_{0k})\right),$$

where the inner value is 0 when $R_n = 0$.

The known-radius selector compares the two nonparametric remainders associated with these constructions. The quantity $u_{n,d}$ is the rare-cell polynomial scale, and $h_{n,d,\sigma}$ is the collision scale that combines the heterogeneity radius with the sparse crossed-cell term.

Algorithm 3 (Selector $\widehat{\tau}_n^\star$)

Given candidate estimators T_n^{poly} and T_n^{col}, each a total measurable map $\mathcal{O}^n \to [-M, M]$, define $\widehat{\tau}_n^\star$ by the following steps.

1. Define the rare-cell polynomial remainder

$$u_{n,d} := \frac{d^2}{n^2 \log^2(en)}.$$

2. Define the collision remainder

$$h_{n,d,\sigma} := \sigma^2 + \frac{d}{n^2}.$$

3. Output the measurable selector

$$\widehat{\tau}_n^\star := \begin{cases} T_n^{\mathrm{col}}, & h_{n,d,\sigma} \le \min\{1, u_{n,d}\}, \\ T_n^{\mathrm{poly}}, & u_{n,d} < \min\{1, h_{n,d,\sigma}\}, \\ 0, & \text{otherwise.} \end{cases}$$

The next result records the two risk inequalities that make the selector possible. The polynomial estimator gives a uniform upper bound with the parametric term and the rare-cell polynomial remainder, while the collision estimator gives a uniform upper bound with the parametric term, the squared radius, and the crossed-cell occupancy remainder.

Theorem 1. [`thm:robust-upper-construction-resolution-all-d`] (All-alphabet upper construction) *For every $0 < \epsilon < 1/2$, there exist a constant $C_\epsilon > 0$ and a polynomial calibration handle $h = (N, \rho)$ with $\rho > 0$ such that, for every $n, d \ge 1$, every $M \ge 1$, and every $0 \le \sigma \le 2$, the following hold:*

- ***(Polynomial calibration.)*** *The handle satisfies $N \leq n$ and $d \leq \rho n \log(en) \Rightarrow 2 \leq K_n$, where K_n is the polynomial degree used by the handle-indexed heavy-light signed one-mark Chebyshev factorial estimator T_n^{poly}.*
- ***(Admissibility.)*** *The estimators T_n^{poly} and T_n^{col}, with T_n^{col} as in Algorithm 2, are measurable maps of the n-sample and take values in $[-M, M]$.*
- ***(Model class.)*** *The risk bounds are uniform over $P \in \mathcal{P}_{d,\epsilon,M,\sigma}$, the model class in Definition 1.*

With $u_{n,d} = d^2/\{n^2 \log^2(en)\}$, the estimators obey

$$\sup_{P \in \mathcal{P}_{d,\epsilon,M,\sigma}} E_{Q_P^{(n)}} \left[\{T_n^{\mathrm{poly}} - \tau(P)\}^2 \right] \leq C_\epsilon M^2 \left\{ \frac{1}{n} + \min(1, u_{n,d}) \right\}$$

and

$$\sup_{P \in \mathcal{P}_{d,\epsilon,M,\sigma}} E_{Q_P^{(n)}} \left[\{T_n^{\mathrm{col}} - \tau(P)\}^2 \right] \leq C_\epsilon M^2 \left(\frac{1}{n} + \sigma^2 + \frac{d}{n^2} \right).$$

Theorem 1 separates the two sources of gain. The polynomial construction is calibrated to the large-alphabet regime through $\log(en)$ and contributes the scale $d^2/\{n^2 \log^2(en)\}$. The collision construction uses the approximate-homogeneity radius: when cell effects are close to their average, missing uncrossed contrasts contribute at the σ^2 scale, and crossed-cell scarcity contributes d/n^2.

Combining the two inequalities with the measurable selector yields the all-alphabet upper benchmark. The selector uses the radius as an input, so the comparison among the three branches is deterministic once n, d, and σ are fixed.

Theorem 2. [`thm:frontier-upper-all-d`] (All-alphabet frontier upper bound) *For every overlap parameter ϵ with $0 < \epsilon < 1/2$, there exist a constant $C_\epsilon > 0$ and a polynomial calibration handle $(N_\epsilon, \rho_\epsilon)$ such that:*

- ***(Handle calibration.)*** *$\rho_\epsilon > 0$, and for every pair of integers $n, d \geq 1$,*

  $$N_\epsilon \leq n, \qquad d \leq \rho_\epsilon n \log(en) \quad \Longrightarrow \quad 2 \leq K_n,$$

  where K_n is the polynomial degree used by the polynomial estimator.
- ***(Index range.)*** *The guarantee applies to every pair of integers $n, d \geq 1$, every $M \geq 1$, and every $0 \leq \sigma \leq 2$.*
- ***(Model class.)*** *The law P ranges over $\mathcal{P}_{d,\epsilon,M,\sigma}$ from Definition 1.*

For these indices, let T_n^{poly} be the admissible polynomial estimator calibrated by $(N_\epsilon, \rho_\epsilon)$, let T_n^{col} be the collision estimator, and let $\widehat{\tau}_n^\star$ be the known-radius selector in Algorithm 3 applied with radius σ to these two estimators, and write

$$r_{n,d,\sigma} = \frac{1}{n} + \min\left\{1, \min\left[\frac{d^2}{n^2 \log^2(en)}, \sigma^2 + \frac{d}{n^2}\right]\right\}$$

for the selector benchmark. Then $\widehat{\tau}_n^\star$ is measurable, satisfies $\widehat{\tau}_n^\star(s) \in [-M, M]$ for every sample s, and

$$\sup_{P \in \mathcal{P}_{d,\epsilon,M,\sigma}} E_{Q_P^{(n)}} \left[(\widehat{\tau}_n^\star - \tau(P))^2 \right] \leq C_\epsilon M^2 r_{n,d,\sigma}.$$

Theorem 2 packages the constructive side of the minimax analysis. For every alphabet size and sample size, the known-radius estimator attains the parametric term plus the smallest available nonparametric remainder among the polynomial, collision, and clipped-zero branches. This all-alphabet statement is the upper half of the two-sided bracket assembled in the next sections.

5 Lower bounds and binary embeddings

The upper bounds in Theorems 1 and 2 are matched against lower bounds obtained by embedding binary hard experiments into the real-outcome model. The source experiments come from the discrete-adjustment constructions of Zeng et al. [2024]. The present section records the embedding map, the transfer of binary converses into the radius-indexed class in Definition 1, and the all-alphabet lower bound that will be combined with the selector bound in the main bracket.

The least-favorable construction uses two binary sources. One source carries an exact-homogeneity collision lower bound, and the other passes through a channel whose attenuation equals one half of the heterogeneity radius. The capped alphabet sizes keep the source experiment within the available d cells for every finite n, d.

The lower-bound proof has a parallel structure. The exact-homogeneity component restricts the real-outcome model to affine images of uniform-mass binary laws with a common cell effect; the binary collision lower bound then gives the baseline $M^2\{n^{-1} + \min(1, d/n^2)\}$. The radius component starts from an arbitrary-mass one-arm binary lower bound, pads the source alphabet into the available d cells, applies a Bernoulli channel with attenuation $\sigma/2$, and rescales outcomes by M. This transport preserves the observed-data comparison up to data processing and turns source target separation into average-treatment-effect separation of order $M\sigma$, giving the term $M^2\sigma^2 \min\{1, u_{n,d}\}$. The exact and radius components are combined with the parametric lower term in the bracket.

Algorithm 4 (Least-favorable handle)

For fixed constants $b_\epsilon^{\mathrm{rad}}, b_\epsilon^{\mathrm{ex}} > 0$ depending only on ϵ, define the two embedded hard families and the channel as follows.

1. Set the channel attenuation
$$\lambda := \frac{\sigma}{2}.$$

2. For the radius component, set the capped alphabet size
$$d_{\mathrm{rad}} := \min\{d, \max(1, \lfloor b_\epsilon^{\mathrm{rad}} n \log(en) \rfloor)\}.$$
Start from a fixed-sample binary one-arm hard family with at most d_{rad} positive-mass cells, and assign zero mass to unused cells.

3. Pass each binary potential response through the hypothesis-independent channel
$$B'(a) \sim \mathrm{Bernoulli}\left\{\frac{1}{2} + \lambda\left(B(a) - \frac{1}{2}\right)\right\},$$

and define the scaled real potential response

$$Y(a) := M\{B'(a) - 1/2\}.$$

4. For the exact-homogeneity component, set the capped alphabet size

$$d_{\mathrm{ex}} := \min\{d, \max(1, \lfloor b_\epsilon^{\mathrm{ex}} n^2 \rfloor)\}.$$

   Start from a uniform-mass binary hard family on at most d_{ex} positive-mass cells, assign zero mass to unused cells, and define

$$Y(a) := M\{B(a) - 1/2\}.$$

5. The construction outputs the radius-channel embedded family, the exact-homogeneity embedded family, and the channel above.

Algorithm 4 fixes the two source-to-target routes used below. The exact-homogeneity route preserves the binary response contrast after the affine scaling by M. The radius route first contracts binary mean differences by the factor $\sigma/2$, so the transported alternatives fit the heterogeneity radius in Assumption 5 while retaining a source separation proportional to $M\sigma$.

The affine embedding is the basic transport map. It leaves the discrete adjustment structure unchanged and rescales binary outcomes to the centered real interval dictated by Assumption 4.

Definition 8. [`synth_4`] (Affine binary-to-real embedding) *For $M \geq 1$, define Φ_M as the affine pushforward that sends a binary d-cell law P to the real-outcome d-cell law obtained from*

$$Y(a) := M\{B(a) - 1/2\}, \qquad Y = Y(A).$$

Under Φ_M, the cell masses and propensities agree with those of P, each positive arm-cell conditional outcome law is pushed forward by $b \mapsto M(b - 1/2)$, and

$$\mu_{ak}(\Phi_M(P)) = M\{\mu_{ak}^{\mathrm{bin}}(P) - 1/2\}, \qquad \tau_k(\Phi_M(P)) = M\{\mu_{1k}^{\mathrm{bin}}(P) - \mu_{0k}^{\mathrm{bin}}(P)\}.$$

The mean-squared-error notation used in the transfer statements is tied to the observed product law in Definition 3 and the average treatment effect functional in Definition 4.

Definition 9. [`synth_7`] (Mean-squared error $\mathrm{MSE}_{P'}(T)$) *For a real-outcome law P' and a total estimator T of the n-sample, define its mean-squared error at P' by*

$$\mathrm{MSE}_{P'}(T) := E_{Q_{P'}^{(n)}}\left[\left(T - \tau(P')\right)^2\right],$$

where $Q_{P'}^{(n)} = (P'_O)^{\otimes n}$ is the observed product law and $\tau(P')$ is the average treatment effect functional.

With these definitions in place, the first transfer result records how the binary classes enter the real-outcome classes and how their risk lower bounds scale. It also states the polynomial upper guarantee attached to the same affine framework, which keeps the subclass comparison and the constructive bound on a common outcome scale.

Proposition 1. [`prop:zeng-class-inclusion-and-lower-transfer`] (Affine subclass transfer) *Let $\mathcal{D}_d^{\mathrm{bin}}(\epsilon)$ and $\mathcal{H}_d^{\mathrm{bin}}(\epsilon)$ be the binary source classes of Definition 7, whose defining conditions are recalled here only through the overlap and homogeneity constraints that this proposition uses:*

$$\mathcal{D}_d^{\mathrm{bin}}(\epsilon) := \{P : \text{ for every positive-mass cell } k,\ \epsilon \le \pi_k(P) \le 1-\epsilon\}.$$

and, for the uniform-mass exactly homogeneous subclass,

$$\mathcal{H}_d^{\mathrm{bin}}(\epsilon) := \left\{ P \in \mathcal{D}_d^{\mathrm{bin}}(\epsilon) : p_k(P) = \frac{1}{d} \text{ for every } k,\ \mu_{1k}^{\mathrm{bin}}(P) - \mu_{0k}^{\mathrm{bin}}(P) = \mu_{1\ell}^{\mathrm{bin}}(P) - \mu_{0\ell}^{\mathrm{bin}}(P) \text{ for all } k, \ell \right\}.$$

For every overlap parameter ϵ with $0 < \epsilon < 1/2$, there are constants $c_{\mathrm{ex}}, c_{\mathrm{rad}}, b_{\mathrm{rad}}, C_{\mathrm{up}} > 0$ and a polynomial calibration handle, consisting of a cutoff $K_0 \in \mathbb{N}$ and an active-range multiplier $A > 0$ such that

$$K_0 \le n, \qquad d \le An\log(en) \qquad \Longrightarrow \qquad 2 \le K_n$$

for all $n, d \ge 1$, where K_n is the polynomial degree of Algorithm 1, with the following property. For every $d \ge 1$ and $M \ge 1$, there is an affine embedding map Φ_M, sending each binary d-cell law to a real-outcome d-cell law,

$$\Phi_M : \{d\text{-cell binary laws}\} \longrightarrow \{d\text{-cell real-outcome laws}\},$$

with the following properties:

- ***(Affine embedding.)*** *For every binary law P, $\Phi_M(P)$ is the affine pushforward $B(a) \mapsto M\{B(a) - 1/2\}$: its observed law is the pushforward of the binary observed law, its cell masses and propensities agree with those of P, each positive arm-cell conditional outcome law is pushed forward by $b \mapsto M(b-1/2)$, each arm-cell mean is M times the corresponding binary mean minus $M/2$, and its full-data law is induced by an affine pushforward of a binary full-data coupling.*

- ***(Unrestricted inclusion.)*** *If $P \in \mathcal{D}_d^{\mathrm{bin}}(\epsilon)$, then $\Phi_M(P) \in \mathcal{P}_{d,\epsilon,M}^{\mathrm{unr}}$, the class in Definition 2.*

- ***(Exact homogeneous inclusion.)*** *If $P \in \mathcal{H}_d^{\mathrm{bin}}(\epsilon)$, then*

  $$\Phi_M(P) \in \mathcal{P}_{d,\epsilon,M,0},$$

  with $\mathcal{P}_{d,\epsilon,M,\sigma}$ as in Definition 1.

- ***(Strict image.)*** *For every $0 \le \sigma \le 2$, there exists a real-outcome law $Q \in \mathcal{P}_{d,\epsilon,M,\sigma}$ that is the image of no binary law:*

  $$\Phi_M(P) \ne Q$$

  for every binary law P.

- ***(Lower transfers and upper estimator.)*** *Write $\mathrm{MSE}_{P'}(T) := E_{Q_{P'}^{(n)}}\left[(T - \tau(P'))^2\right]$ for the mean-squared error of a total estimator T under the n-sample law of a real-outcome law P'. For every $n \ge 1$ and $0 \le \sigma \le 2$, if $d \le c_{\mathrm{ex}} n^2$, then every total measurable estimator with values in $[-M, M]$ has some $P \in \mathcal{H}_d^{\mathrm{bin}}(\epsilon)$ satisfying*

  $$c_{\mathrm{ex}} M^2 \left(\frac{1}{n} + \frac{d}{n^2}\right) \le \mathrm{MSE}_{\Phi_M(P)}(\widehat{\tau}),$$

and the minimax risk in Definition 5 satisfies

$$c_{\mathrm{ex}} M^2 \left(\frac{1}{n} + \frac{d}{n^2} \right) \leq \mathsf{R}_{n,d,\epsilon,M,\sigma}.$$

For all $n \geq 1$ and $0 \leq \sigma \leq 2$, the radius-channel transfer gives

$$c_{\mathrm{rad}} M^2 \sigma^2 \min\{1, u_{n,d}\} \leq \mathsf{R}_{n,d,\epsilon,M,\sigma},$$

together with a transfer certificate with constants c_{rad} and b_{rad}. The estimator T_n^{poly} determined by the polynomial calibration handle is measurable, takes values in $[-M, M]$, and for every $Q \in \mathcal{P}_{d,\epsilon,M,\sigma}$,

$$\mathrm{MSE}_Q(T_n^{\mathrm{poly}}) \leq C_{\mathrm{up}} M^2 \left(\frac{1}{n} + \min\{1, u_{n,d}\} \right).$$

Proposition 1 makes the binary-to-real comparison precise. The binary arbitrary-mass source class embeds into the unrestricted real-outcome class, while the uniform-mass exactly homogeneous source embeds into the zero-radius real-outcome class. The transferred exact-homogeneity lower bound contributes the $M^2(n^{-1} + d/n^2)$ scale on its active alphabet range, and the radius-channel transfer contributes the $M^2\sigma^2 \min\{1, u_{n,d}\}$ scale for the rare-cell component. The final clause records that the polynomial estimator applies over the larger real-outcome class in Definition 1, so the embedding is used for converses while the constructive guarantee remains same-class.

The next statement extends the exact-homogeneity lower transfer to all alphabet sizes by saturating the collision term at one. It is the exact-radius component of the lower benchmark used in the bracket.

Lemma 2. [`lem:scaled-binary-exact-lower-transfer-all-d`] (Scaled binary lower transfer) *For every overlap parameter ϵ with $0 < \epsilon < 1/2$, there exists a constant $c_\epsilon > 0$ such that the following statement holds. For every $n, d \in \mathbb{N}$, $M \in \mathbb{R}$, and $\sigma \in \mathbb{R}$ satisfying*

- ***(Sample and alphabet.)*** *$n \geq 1$ and $d \geq 1$;*
- ***(Outcome scale.)*** *$M \geq 1$;*
- ***(Radius range.)*** *$0 \leq \sigma \leq 2$,*

the minimax mean-squared risk $\mathsf{R}_{n,d,\epsilon,M,\sigma}$ from Definition 5 obeys

$$c_\epsilon M^2 \left\{ \frac{1}{n} + \min\left(1, \frac{d}{n^2}\right) \right\} \leq \mathsf{R}_{n,d,\epsilon,M,\sigma}.$$

The proof is deferred to Section E.

Lemma 2 supplies the lower-bound baseline that persists at every radius because exact homogeneity is included within each radius-indexed class. The term n^{-1} is the parametric sampling contribution, and the capped d/n^2 term is the collision contribution inherited from the binary exactly homogeneous source experiment of Zeng et al. [2024] after scaling by M^2.

The radius-channel converse adds the rare-cell component governed by the heterogeneity radius. The construction in Algorithm 4 converts a binary source separation for $\psi(P)$, the binary average response contrast introduced in Definition 6, into a real-outcome separation for $\tau(P)$ through the displayed affine channel.

Theorem 3. [`thm:radius-channel-converse-all-d`] (All-d radius converse) *For every $0 < \epsilon < 1/2$, there exist constants c_ϵ and $b_\epsilon^{\mathrm{rad}}$ such that*

$$0 < c_\epsilon \le 1, \qquad 0 < b_\epsilon^{\mathrm{rad}}.$$

For all integers $n, d \ge 1$, all $M \ge 1$, and all $0 \le \sigma \le 2$, the minimax mean-squared risk $\mathsf{R}_{n,d,\epsilon,M,\sigma}$ of Definition 5 satisfies

$$c_\epsilon M^2 \left\{ \frac{1}{n} + \min\left(1, \frac{d}{n^2}\right) + \sigma^2 \min\left(1, \frac{d^2}{n^2 \log^2(en)}\right) \right\} \le \mathsf{R}_{n,d,\epsilon,M,\sigma}.$$

Moreover, for the same constants and every such n, d, M, σ, there are a capped alphabet size d_{rad}, a source family, a padding map into $\{1, \ldots, d\}$, an embedding into $\mathcal{P}_{d,\epsilon,M,\sigma}$, and a coupling realizing the radius-channel construction with source constant $b_\epsilon^{\mathrm{rad}}$. Along this realization, for all source laws P_0, P_1,

$$\tau(P_1) - \tau(P_0) = M\frac{\sigma}{2}\{\psi(P_1) - \psi(P_0)\},$$

and for every estimator $\widehat{\tau}$ taking values in $[-M, M]$, some source law P satisfies

$$c_\epsilon M^2 \sigma^2 \min\left(1, \frac{d^2}{n^2 \log^2(en)}\right) \le \mathbb{E}_P\{(\widehat{\tau} - \tau(P))^2\}.$$

Theorem 3 establishes the all-alphabet lower benchmark for the same minimax risk studied by the selector upper bound. Its first two terms are the exact-homogeneity baseline from Lemma 2; the third term is the radius-sensitive rare-cell term generated by the channel. The displayed coupling identity shows why the scale is $M^2\sigma^2$: the affine map contributes the factor M, the channel contributes the factor $\sigma/2$, and mean-squared risk squares the resulting target separation. The capped construction keeps the embedded source inside $\mathcal{P}_{d,\epsilon,M,\sigma}$ for every finite alphabet size, so the lower bound is available at the same n, d, M, σ indices as the upper result in Theorem 2.

Together, Lemma 2 and Theorem 3 provide the converse side of the paper's minimax bracket. The next section combines these lower bounds with the selector guarantee in Theorem 2 and then characterizes the regimes in which the two benchmarks agree in order.

6 Main bracket and matched regimes

The lower bounds in Lemma 2 and Theorem 3 and the selector upper bound in Theorem 2 combine into the paper's main finite-sample statement. The result is indexed by the same sample size, alphabet size, outcome scale, overlap constant, and heterogeneity radius as the model class in Definition 1; the constants depend only on the fixed overlap parameter. In this section, the all-alphabet lower benchmark is the capped converse benchmark $\ell^{\mathrm{cap}}_{n,d,\sigma}$, while the triangular algebra later uses $\ell^{\mathrm{tri}}_{n,d,\sigma} = b_{n,d} + \sigma^2 u_{n,d}$ for the product form. The two displayed benchmarks separate the exact-homogeneity collision baseline from the radius-sensitive rare-cell component on the lower side, and the best available selector branch on the upper side.

Theorem 4. [`thm:two-sided-minimax-bracket-all-d`] (All-alphabet minimax bracket) *For every overlap parameter ϵ with $0 < \epsilon < 1/2$, there are real constants c_ϵ and C_ϵ such that*

$$0 < c_\epsilon \le 1, \qquad 1 \le C_\epsilon, \qquad c_\epsilon \le C_\epsilon.$$

For every $n, d \in \mathbb{N}$, $M \in \mathbb{R}$, and $\sigma \in \mathbb{R}$, suppose that

- ***(Sample and alphabet sizes.)*** *$n \geq 1$ and $d \geq 1$.*
- ***(Outcome envelope.)*** *$M \geq 1$.*
- ***(Radius range.)*** *$0 \leq \sigma \leq 2$.*

Let $\mathsf{R}_{n,d,\epsilon,M,\sigma}$ be the minimax mean-squared risk in Definition 5 over the real-outcome class $\mathcal{P}_{d,\epsilon,M,\sigma}$ with fixed overlap ϵ, conditional exchangeability, consistency, conditional mean normalization, conditional second central moment envelope M, and approximate-homogeneity radius σM. Define

$$u_{n,d} = \frac{d^2}{n^2 \log^2(en)}, \qquad h_{n,d,\sigma} = \sigma^2 + \frac{d}{n^2},$$

and write

$$\ell_{n,d,\sigma} = \frac{1}{n} + \min\left\{1, \frac{d}{n^2}\right\} + \sigma^2 \min\left\{1, u_{n,d}\right\}, \qquad r_{n,d,\sigma} = \frac{1}{n} + \min\left\{1, u_{n,d}, h_{n,d,\sigma}\right\}.$$

Then

$$c_\epsilon M^2 \ell_{n,d,\sigma} \leq \mathsf{R}_{n,d,\epsilon,M,\sigma} \leq C_\epsilon M^2 r_{n,d,\sigma}.$$

Theorem 4 gives a same-class minimax bracket for the scalar average treatment effect over arbitrary cell masses and real outcomes under the moment envelope in Definition 1. The upper benchmark is constructive through the known-radius selector in Algorithm 3. The lower benchmark $\ell^{\mathrm{cap}}_{n,d,\sigma}$ adds two sources of difficulty: the capped collision term already present at exact homogeneity and the radius-channel rare-cell term transported in Theorem 3. Thus the radius governs whether borrowing across cells, crossed-cell contrasts, or clipping to the outcome scale supplies the sharpest available upper benchmark.

The endpoint reductions turn the bracket into familiar rates at the two ends of the radius scale. At $\sigma = 0$, the collision remainder d/n^2 is the relevant sparse-cell term. At radius two, Lemma 1 identifies the radius-indexed class with the unrestricted-radius class, and the polynomial remainder gives the unrestricted all-alphabet benchmark associated with large-alphabet functional estimation [Jiao et al., 2015, Wu and Yang, 2016].

Proposition 2. [`prop:endpoint-reductions-all-d`] (Endpoint rate reductions) *There exist constants $c, C \in (0, \infty)$, with $c \leq C$, such that the following hold.*

- ***(Exact homogeneity.)*** *For every pair of integers $n, d \geq 1$,*

$$c\left\{n^{-1} + \min\left(1, \frac{d}{n^2}\right)\right\} \leq n^{-1} + \min\left\{1, \frac{d^2}{n^2\log^2(en)}, \frac{d}{n^2}\right\} \leq C\left\{n^{-1} + \min\left(1, \frac{d}{n^2}\right)\right\},$$

and

$$c\left\{n^{-1} + \min\left(1, \frac{d}{n^2}\right)\right\} \leq n^{-1} + \min\left(1, \frac{d}{n^2}\right) \leq C\left\{n^{-1} + \min\left(1, \frac{d}{n^2}\right)\right\}.$$

- ***(Unrestricted radius.)*** *For every pair of integers $n, d \geq 1$,*

$$c\left\{n^{-1} + \min\left(1, \frac{d^2}{n^2\log^2(en)}\right)\right\} \leq n^{-1} + \min\left(1, \frac{d^2}{n^2\log^2(en)}\right) \leq C\left\{n^{-1} + \min\left(1, \frac{d^2}{n^2\log^2(en)}\right)\right\},$$

and

$$c\left\{n^{-1} + \min\left(1, \frac{d^2}{n^2\log^2(en)}\right)\right\} \leq n^{-1} + \min\left(1, \frac{d}{n^2}\right) + 4\min\left(1, \frac{d^2}{n^2\log^2(en)}\right) \leq C\left\{n^{-1} + \min\left(1, \frac{d^2}{n^2\log^2(en)}\right)\right\}$$

- ***(Class equality at radius two.)*** *For every alphabet size d, every $\epsilon \in \mathbb{R}$, and every $M \in \mathbb{R}$, the laws represented by $\mathcal{P}^{\mathrm{unr}}_{d,\epsilon,M}$ in Definition 2 are exactly the laws represented by $\mathcal{P}_{d,\epsilon,M,2}$ in Definition 1.*

- ***(Endpoint identity.)*** *For every pair of integers $n, d \geq 1$, the radius-two upper frontier expression is exactly*
$$n^{-1} + \min\left(1, \frac{d^2}{n^2 \log^2(en)}\right).$$

Proposition 2 calibrates the two endpoints of the regime summary. Exact homogeneity gives the rate $n^{-1} + \min(1, d/n^2)$, matching the collision lower transfer from Zeng et al. [2024]. The unrestricted endpoint gives $n^{-1} + \min\{1, d^2/[n^2 \log^2(en)]\}$, the polynomial large-alphabet scale. The class equality clause also fixes the interpretation of the radius normalization: radius two represents the full unrestricted-radius law class introduced in Definition 2.

The regime theorem supplies the order comparisons away from the endpoints. It covers every fixed positive radius, the saturation region, small alphabets, and the two parametric-dominance elbows. It also records the residual shrinking-radius wedge generated by the proved benchmarks. The symbol $\ell_{n,d,\sigma}$ in the theorem display below is local to that anchored statement and denotes the triangular product benchmark; in the surrounding prose it is written as $\ell^{\mathrm{tri}}_{n,d,\sigma}$ to distinguish it from the capped converse benchmark in Theorem 4.

Theorem 5. [`thm:fixed-interior-tightness-and-shrinking-radius-gap-all-d`] (Fixed-radius wedge characterization) *Fix $0 < \epsilon < 1/2$. For $n, d \in \mathbb{N}$, put $L_n = \log(en)$,*

$$b_{n,d} = \frac{1}{n} + \frac{d}{n^2}, \qquad u_{n,d} = \frac{d^2}{n^2 L_n^2}, \qquad h_{n,d,\sigma} = \sigma^2 + \frac{d}{n^2}, \qquad r_{n,d,\sigma} = \frac{1}{n} + \min\{1, \min(u_{n,d}, h_{n,d,\sigma})\},$$

and define
$$q_{n,d,\sigma} = \frac{1}{n} + \min\left\{u_{n,d}, \sigma^2 + \frac{d}{n^2}\right\}, \qquad \ell_{n,d,\sigma} = b_{n,d} + \sigma^2 u_{n,d}.$$

Let $\mathsf{R}_{n,d,\epsilon,M,\sigma}$ be the minimax mean-squared risk in Definition 5. Then all of the following hold.

- ***(Triangular benchmarks.)*** *For every $n, d \in \mathbb{N}$ and every $0 \leq \sigma \leq 2$, if*
$$L_n^2 < d < nL_n,$$
then
$$\frac{1}{2}\{b_{n,d} + \min(u_{n,d}, \sigma^2)\} \leq q_{n,d,\sigma} \leq b_{n,d} + \min(u_{n,d}, \sigma^2), \qquad \ell_{n,d,\sigma} = b_{n,d} + \sigma^2 u_{n,d}.$$

- ***(Fixed interior radii.)*** *For every $\underline{\sigma} \in (0, 2]$, there exist constants c, C with $0 < c \leq C$ such that, for every $n, d \in \mathbb{N}$, every $M \geq 1$, and every σ satisfying $\underline{\sigma} \leq \sigma \leq 2$,*
$$cM^2 r_{n,d,\sigma} \leq \mathsf{R}_{n,d,\epsilon,M,\sigma} \leq CM^2 r_{n,d,\sigma},$$
whenever $n \geq 1$ and $d \geq 1$.

- ***(Matching elbows.)*** *For every $K \geq 0$, there exist constants c, C with $0 < c \leq C$ such that, for every $n, d \in \mathbb{N}$, every $M \geq 1$, and every $0 \leq \sigma \leq 2$,*

$$cM^2 r_{n,d,\sigma} \leq \mathsf{R}_{n,d,\epsilon,M,\sigma} \leq CM^2 r_{n,d,\sigma},$$

whenever $n \geq 1$, $d \geq 1$, and at least one of

$$1 \leq u_{n,d}, \qquad u_{n,d} \leq Kb_{n,d}, \qquad \sigma^2 \leq Kb_{n,d}$$

holds.

- ***(Exact algebraic elbow.)*** *For every $n, d \in \mathbb{N}$ with $n \geq 1$,*

$$u_{n,d} \leq b_{n,d} \quad \Longleftrightarrow \quad d^2 \leq nL_n^2 + dL_n^2.$$

- ***(Sufficient elbows.)*** *For every $n, d \in \mathbb{N}$ with $n \geq 1$,*

$$d \leq \sqrt{n}\, L_n \quad \Longrightarrow \quad u_{n,d} \leq b_{n,d},$$

and

$$d \leq L_n^2 \quad \Longrightarrow \quad u_{n,d} \leq b_{n,d}.$$

- ***(Residual wedge implication.)*** *For any sequences $d_n \in \mathbb{N}$ and $\sigma_n \in \mathbb{R}$, if eventually $d_n \geq 1$ and $0 \leq \sigma_n \leq 2$, and if the frontier-to-capped-converse benchmark ratio*

$$\frac{r_{n,d_n,\sigma_n}}{n^{-1} + \min\{1, d_n/n^2\} + \sigma_n^2 \min\{1, u_{n,d_n}\}} \to \infty,$$

then the sequence lies in the residual wedge:

$$\frac{b_{n,d_n}}{u_{n,d_n}} \to 0, \qquad u_{n,d_n} \to 0, \qquad \frac{b_{n,d_n}}{\sigma_n^2} \to 0, \qquad \sigma_n^2 \to 0.$$

- ***(Nonempty wedge.)*** *The sequence $d_n = n$ and $\sigma_n = L_n^{-1/2}$ lies in the residual wedge.*

- ***(Diagonal divergence.)*** *Along $d_n = n$ and $\sigma_n = L_n^{-1/2}$,*

$$\frac{q_{n,n,L_n^{-1/2}}}{\ell_{n,n,L_n^{-1/2}}} \asymp L_n$$

eventually, and

$$\frac{q_{n,n,L_n^{-1/2}}}{\ell_{n,n,L_n^{-1/2}}} \to \infty.$$

Theorem 5 gives the rate consequences of the bracket. For every radius bounded away from zero, the minimax risk is equivalent to the selector benchmark $M^2 r_{n,d,\sigma}$. The same equivalence holds in the saturated region $1 \leq u_{n,d}$, in regimes where the polynomial term is dominated by the exact-homogeneity baseline, and in regimes where the squared radius is dominated by that baseline. The algebraic and sufficient elbow clauses identify concrete alphabet ranges, including $d \leq \sqrt{n} \log(en)$ and $d \leq \log^2(en)$, where the baseline already controls the polynomial remainder.

Regime	Condition	Matched benchmark or localized separation
Exact homogeneity	$\sigma = 0$	$M^2\{n^{-1} + \min(1, d/n^2)\}$
Unrestricted radius	$\sigma = 2$	$M^2\{n^{-1} + \min(1, u_{n,d})\}$
Fixed positive radius	$\underline{\sigma} \le \sigma \le 2$	$M^2 r_{n,d,\sigma}$
Saturation and elbows	$1 \le u_{n,d}$, or $u_{n,d} \le K b_{n,d}$, or $\sigma^2 \le K b_{n,d}$	$M^2 r_{n,d,\sigma}$
Selector/converse benchmark gap	$b_{n,d} \ll u_{n,d} \ll 1$ and $b_{n,d} \ll \sigma^2 \ll 1$	selector benchmark $M^2\{b_{n,d} + \min(u_{n,d}, \sigma^2)\}$; product-form converse benchmark $M^2\{b_{n,d} + \sigma^2 u_{n,d}\}$

Table 1: Matched-regime summary for the two-sided minimax bracket. The final row summarizes the proved comparison between displayed selector and converse benchmarks in the residual wedge; exact minimax-rate conclusions are the rows where the benchmark is listed as matched.

Table 1 summarizes the rate statements across the regimes of the (d, σ) scale; boundary constants and logarithmic factors are governed by the displayed formulas in Theorem 5.

The final row of Table 1 is a localization statement about the displayed benchmarks. In the intermediate alphabet range $L_n^2 < d < nL_n$, the selector benchmark is comparable to $b_{n,d} + \min(u_{n,d}, \sigma^2)$, while the product-form converse benchmark is $\ell^{\mathrm{tri}}_{n,d,\sigma} = b_{n,d} + \sigma^2 u_{n,d}$. These expressions agree in the fixed-radius and parametric-dominance regimes described in the theorem. The residual shrinking-radius wedge consists of sequences where the baseline is small relative to both the polynomial scale and the squared radius, while both the polynomial scale and squared radius vanish. The diagonal sequence $d_n = n$, $\sigma_n = L_n^{-1/2}$ shows that this localized separation region is nonempty and that the ratio between the displayed upper and lower benchmarks can diverge at logarithmic order.

Taken together, Theorems 4 and 5 and Proposition 2 provide the all-alphabet minimax summary. The paper establishes a constructive upper benchmark and a same-class lower benchmark for the radius-indexed real-outcome model, matches them at the two endpoints and throughout the fixed-positive-radius and elbow regimes, and isolates the shrinking-radius intermediate region generated by the proved finite-sample inequalities.

6.1 Binary collision comparison

A useful secondary comparison is with the published binary collision estimator of Zeng et al. [2024]. The following statement keeps that comparison separate from the same-class minimax bracket: it assumes the published binary collision guarantee and then records the algebraic relationship between the present selector benchmark and the collision remainder.

Proposition 3. [`thm:published-binary-collision-comparison`] (Binary collision comparison) *Fix $\epsilon \in (0, 1/2)$. For a binary d-cell law P write P_n for its n-sample product experiment and s for the published binary occupancy-weighted collision estimator, and say that P has binary maximal heterogeneity at radius σ_{bin} when $\sigma_{\mathrm{bin}} \ge 0$, every cell satisfies $\left|\mu^{\mathrm{bin}}_{1k}(P) - \mu^{\mathrm{bin}}_{0k}(P) - \tau(P)\right| \le \sigma_{\mathrm{bin}}$, and some cell attains equality. Suppose the published binary collision guarantee holds at overlap ϵ: there are constants $C_\epsilon, b_\epsilon > 0$ such that, for every $n, d \in \mathbb{N}$ with $d > 0$, every binary* *law P with overlap ϵ, and every $\sigma_{\text{bin}} \in \mathbb{R}$ satisfying binary maximal heterogeneity at radius σ_{bin},*

$$|E_{P_n}[s] - \tau(P)| \leq \sigma_{\text{bin}} + 2\exp\left\{-b_\epsilon \frac{n^2}{\max\{n,d\}}\right\}$$

and

$$\operatorname{Var}_{P_n}(s) \leq C_\epsilon\left(\sigma_{\text{bin}}^2 + \frac{d}{n^2} + \frac{1}{n}\right).$$

Define

$$u_{n,d} := \frac{d^2}{n^2\log^2(en)} \qquad \text{and} \qquad h_{n,d,\sigma} := \sigma^2 + \frac{d}{n^2}.$$

Then the published binary collision guarantee above and the following algebraic comparison hold simultaneously:

- ***(Pointwise comparison.)*** *For every $n, d \in \mathbb{N}$ and $\sigma \in \mathbb{R}$ such that $n > 0$, $d > 0$, and $0 \leq \sigma \leq 2$,*

$$\min\{1, \min\{u_{n,d}, h_{n,d,\sigma}\}\} \leq h_{n,d,\sigma}.$$

- ***(Asymptotic comparison.)*** *For every pair of sequences $d_n \in \mathbb{N}$ and $\sigma_n \in \mathbb{R}$ satisfying $d_n > 0$ and $0 \leq \sigma_n \leq 2$ for all $n > 0$, if*

$$\frac{u_{n,d_n}}{h_{n,d_n,\sigma_n}} \to 0 \qquad \text{and} \qquad \frac{d_n}{n\log(en)} \to 0,$$

then

$$\frac{\min\{1, \min\{u_{n,d_n}, h_{n,d_n,\sigma_n}\}\}}{h_{n,d_n,\sigma_n}} \to 0.$$

Proposition 3 has two roles. First, it states the exact binary occupancy guarantee used for the comparison, including its bias and variance scales. Second, it shows that the selector's nonparametric term is bounded pointwise by the collision remainder, and that it is asymptotically smaller along sequences where the polynomial scale is negligible relative to the collision scale and the alphabet lies below $n\log(en)$ in the displayed sense. This makes the comparison presentation-level: the headline bracket in Theorem 4 rests on the same-class upper and lower statements assembled in Theorems 2 and 3.

7 Discussion and limitations

The bracket in Theorem 4 gives a finite-sample interpretation of the known-radius problem on a single real-outcome observational class. The estimator side is entirely constructive through $\widehat{\tau}_n^\star$ in Algorithm 3, and the converse side is stated for the same class $\mathcal{P}_{d,\epsilon,M,\sigma}$ from Definition 1. The radius σM is an intrinsic coordinate of that class: it is the maximal supported-cell effect deviation inside the same fixed-overlap, arbitrary-cell-mass model used to define the minimax risk in Definition 5.

The bracket benchmarks also give a concrete reading of the two estimation mechanisms. When the radius is small enough for crossed cells to be informative, the collision branch uses observed treated-control meetings within cells. When the alphabet is sparse enough for rare-cell approximation to dominate, the polynomial branch uses the heavy-light construction in Algorithm 1. The selector benchmark in Theorem 2 records the better of these two mechanisms, together with the clipped-zero branch, while Theorem 5 identifies the endpoint, fixed-positive-radius, saturation, and parametric-dominance regimes where the bracket is order matched.

7.1 Future work

The matched-regime summary in Theorem 5 leaves a sharply described residual shrinking-radius wedge for further development. The following remark records the target problem in that region using the same notation as the regime theorem.

Remark 1. [`oeq:shrinking-radius-frontier`] (Shrinking-radius frontier) *Let*

$$L := \log(en), \qquad b := n^{-1} + \frac{d}{n^2}, \qquad u := \frac{d^2}{n^2 L^2}.$$

In the residual shrinking-radius wedge

$$b \ll u \ll 1, \qquad b \ll \sigma^2 \ll 1,$$

equivalently up to boundary constants

$$\sqrt{n}\, L \ll d \ll nL, \qquad \sigma^2 \gg b,$$

a natural next question is whether one can either construct a realization-wise radius-constrained paired-cell moment-matching fuzzy experiment in $\mathcal{P}_{d,\epsilon,M,\sigma}$ with ATE separation of order

$$M \min\left\{\sigma, \frac{d}{nL}\right\}$$

and sample-mixture total variation bounded away from one, or construct an explicit total estimator whose risk is strictly smaller in order than the selector benchmark

$$M^2\{b + \min(u, \sigma^2)\},$$

with the product benchmark

$$M^2(b + \sigma^2 u)$$

as the target rate. The hypothesis-independent channel yields the product separation

$$M\sigma \min\left\{1, \frac{d}{nL}\right\}.$$

Remark 1 records two possible mathematical routes in the residual shrinking-radius wedge. One route asks for a radius-constrained least-favorable construction with ATE separation at the selector scale; the other asks for an estimator whose risk improves on the current selector benchmark toward the product scale suggested by the channel lower bound. The divergence example in Theorem 5 places $d_n = n$ and $\sigma_n = \log(en)^{-1/2}$ inside this region and gives logarithmic separation between the displayed selector and converse benchmarks.

7.2 Limitations

The results are scoped to fixed overlap, known outcome scale M, known heterogeneity radius σ, and scalar average treatment effect estimation. Within that scope, Theorem 4 gives risk bounds for clipped estimators in $\mathcal{T}_{n,M}$ from Definition 5, and Theorem 2 constructs a selector that uses the radius as an input. The inference target is mean-squared error for point estimation over $\mathcal{P}_{d,\epsilon,M,\sigma}$, so confidence intervals, adaptive radius selection, and data-driven choice of M are directions beyond the estimation bracket established here.

The moment and overlap assumptions also define the statistical environment in which the bracket should be read. Fixed overlap in Assumption 3 keeps treatment and control assignment probabilities uniformly away from zero and one in every supported cell. Mean normalization and the second central moment envelope in Assumptions 4 and 6 set the real-outcome scale under weak conditional moment control. Approximate homogeneity in Assumption 5 supplies the radius that determines which estimator branch is selected. Under exactly these conditions, the paper characterizes the delivered finite-sample minimax bracket and its matched regimes for arbitrary cell masses.

Appendices

A Binary source lemmas

The lower-bound transfers in Proposition 1, Lemma 2, and Theorem 3 use binary discrete-adjustment experiments, originating in Zeng et al. [2024], as source problems. This appendix records the binary source lemmas in the notation used by the main text, with attribution and statistical context supplied by the citation. The first definition fixes the fixed-sample exactly homogeneous binary experiment and its binary average response contrast.

Definition 10. [`synth_3`] (Fixed-sample binary homogeneous experiment) *For integers $n, d \geq 1$ and $0 < \epsilon < 1/2$, let $\mathcal{H}^{\mathrm{bin}}_{n,d}(\epsilon)$ be the fixed-sample binary-outcome exactly homogeneous source experiment consisting of the product laws $P^{\otimes n}$ generated by laws $P \in \mathcal{H}^{\mathrm{bin}}_{d}(\epsilon)$. Its binary average response contrast is*

$$\psi(P) = \frac{1}{d}\sum_{k=1}^{d}\{\mu^{\mathrm{bin}}_{1k}(P) - \mu^{\mathrm{bin}}_{0k}(P)\}.$$

Definition 10 packages the source experiment used by the exact-homogeneity collision transfer. It aligns the fixed-sample product laws with the binary source classes introduced in Definition 7 and gives the target $\psi(P)$ that is later scaled into the real-outcome average treatment effect through the affine embeddings.

The first finite-sample collision source lemma concerns the exactly homogeneous binary source experiment. It supplies the d/n^2 component used in the exact-radius part of the real-outcome lower benchmark.

Lemma 3. [`lem:binary-exact-collision-lower`] (Binary collision lower bound) *Let $u_\epsilon := \min\{1/2,\, 8(1/2-\epsilon)^2\}$. Suppose that:*

- ***(Sample and alphabet.)*** *$n, d \in \mathbb{N}$ satisfy $n \geq 1$ and $d \geq 1$.*
- ***(Overlap.)*** *The overlap parameter $\epsilon \in \mathbb{R}$ satisfies $0 < \epsilon < 1/2$.*
- ***(Collision regime.)*** *The dimensions satisfy $d \leq n^2$.*

For the fixed-sample binary exactly homogeneous source experiment $\mathcal{H}^{\mathrm{bin}}_{n,d}(\epsilon)$, generated by binary laws with uniform cell masses, propensities in $[\epsilon, 1-\epsilon]$, and a common cell effect across all *cells, the minimax mean-squared risk for estimating the binary average response contrast $\psi(P)$ satisfies*

$$\frac{u_\epsilon^2}{64}\frac{d}{n^2} \le \inf_{\widehat{\psi}} \sup_{P\in\mathcal{H}^{\mathrm{bin}}_{n,d}(\epsilon)} E_{P^{\otimes n}}\left[\{\widehat{\psi}-\psi(P)\}^2\right],$$

where the infimum ranges over all measurable real-valued estimators $\widehat{\psi}$ of the n-sample.

Proof of Lemma 3. Set

$$u_\epsilon := \min\{1/2, 8(1/2-\epsilon)^2\}, \qquad \Delta := u_\epsilon \frac{\sqrt{d}}{n}, \qquad \rho := \sqrt{\Delta/8}.$$

Write $\eta := 1/2-\epsilon$. From $0<\epsilon<1/2$, $d\le n^2$, and $n,d\ge 1$,

$$0<u_\epsilon, \qquad u_\epsilon \le \frac{1}{2}, \qquad u_\epsilon \le 8\eta^2, \qquad \sqrt{d}\le n.$$

Hence

$$0\le \Delta \le u_\epsilon \le \frac{1}{2}, \qquad \rho^2=\frac{\Delta}{8}, \qquad \rho\le\eta, \qquad \rho\le\frac{1}{4}, \qquad \rho<\frac{1}{2}, \qquad \rho+\frac{\Delta}{4}\le\frac{1}{2}.$$

These are the admissibility inequalities used below: $\rho\le 1/2-\epsilon$ gives overlap, while $\rho<1/2$ and $\rho+\Delta/4\le 1/2$ keep the Bernoulli means in $[0,1]$.

For each sign vector $\xi\in\{\pm1\}^d$, define a uniform-mass binary law P_ξ with

$$\pi_k^{(\xi)} = \frac{1}{2}+\rho\xi_k, \qquad \mu_{1k}^{(\xi)} = \frac{1}{2}-\rho\xi_k+\frac{\Delta}{4}, \qquad \mu_{0k}^{(\xi)} = \frac{1}{2}-\rho\xi_k-\frac{\Delta}{4}.$$

Let P_0 be the corresponding centered law with $\Delta=\rho=0$. The displayed inequalities make each P_ξ and P_0 members of the fixed-sample binary exactly homogeneous source experiment in Definition 10, and

$$\psi(P_\xi)=\frac{\Delta}{2}, \qquad \psi(P_0)=0.$$

Let $Q_0=P_0^{\otimes n}$ and let $\overline{Q}=2^{-d}\sum_\xi P_\xi^{\otimes n}$ be the uniform mixture over signs. For $\xi,\xi'\in\{\pm1\}^d$, put $S(\xi,\xi') := \sum_{k=1}^d \xi_k\xi_k'$ and

$$\Gamma := \Delta-\frac{\Delta^2}{2}+\frac{\Delta^3}{8}.$$

The centered one-observation mass satisfies $p_0(k,a,y)=1/(4d)$. For a fixed cell k, write $s=\xi_k$ and $t=\xi_k'$. Summing over the two arms and two outcomes gives

$$\begin{aligned}
&\sum_{a,y}\frac{p_\xi(k,a,y)p_{\xi'}(k,a,y)}{p_0(k,a,y)}\\
&=\frac{4}{d}\Bigg[\left(\frac{1}{2}+\rho s\right)\left(\frac{1}{2}+\rho t\right)\left\{\left(\frac{1}{2}-\rho s+\frac{\Delta}{4}\right)\left(\frac{1}{2}-\rho t+\frac{\Delta}{4}\right)+\left(\frac{1}{2}+\rho s-\frac{\Delta}{4}\right)\left(\frac{1}{2}+\rho t-\frac{\Delta}{4}\right)\right\}\\
&\qquad+\left(\frac{1}{2}-\rho s\right)\left(\frac{1}{2}-\rho t\right)\left\{\left(\frac{1}{2}-\rho s-\frac{\Delta}{4}\right)\left(\frac{1}{2}-\rho t-\frac{\Delta}{4}\right)+\left(\frac{1}{2}+\rho s+\frac{\Delta}{4}\right)\left(\frac{1}{2}+\rho t+\frac{\Delta}{4}\right)\right\}\Bigg].
\end{aligned}$$

Using $\rho^2 = \Delta/8$, this expression reduces in the two sign cases to

$$\frac{1}{d}\begin{cases} 1+\Delta-\Delta^2/2+\Delta^3/8, & st=1, \\ 1-\Delta+\Delta^2/2-\Delta^3/8, & st=-1, \end{cases} = \frac{1}{d}\{1+\Gamma st\}.$$

Summing over cells gives the one-observation overlap identity

$$\sum_z \frac{p_\xi(z)p_{\xi'}(z)}{p_0(z)} = 1+\frac{\Gamma}{d}S(\xi,\xi').$$

Therefore the finite product second-moment identity gives

$$1+\chi^2(\overline{Q}\|Q_0) = 2^{-2d}\sum_{\xi,\xi'}\left(1+\frac{\Gamma}{d}S(\xi,\xi')\right)^n = 2^{-2d}\sum_{\xi,\xi'}\left(1+\frac{2(\Gamma/2)}{d}S(\xi,\xi')\right)^n.$$

Since $0 \le \Delta \le 1/2$,

$$0 \le \Gamma \le \Delta \le 1/2.$$

We now prove the finite-sign average bound used at $\gamma = \Gamma/2$. Let $\gamma \ge 0$, $2\gamma \le 1$, and $2n^2\gamma^2 \le d\log 2$. With $c := 2\gamma/d$, every sign-pair sum S lies in $[-d,d]$, so $cS \ge -1$. Hence $(1+cS)^n \le \exp(ncS)$, and

$$\begin{aligned} 2^{-2d}\sum_{\xi,\xi'}(1+cS(\xi,\xi'))^n &\le 2^{-2d}\sum_{\xi,\xi'}\exp\{ncS(\xi,\xi')\} \\ &= \prod_{k=1}^d\left(\frac{1}{4}\sum_{s,t\in\{\pm1\}}\exp(ncst)\right) \\ &= \{\cosh(nc)\}^d \le \exp\left\{\frac{dn^2c^2}{2}\right\} = \exp\left\{\frac{2n^2\gamma^2}{d}\right\} \le 2. \end{aligned}$$

For $\gamma = \Gamma/2$, the remaining regularity condition follows from

$$2n^2(\Gamma/2)^2 = \frac{n^2\Gamma^2}{2} \le \frac{n^2\Delta^2}{2} = \frac{u_\epsilon^2 d}{2} \le \frac{d}{8} \le d\log 2,$$

using $\Delta^2 = u_\epsilon^2 d/n^2$, $u_\epsilon \le 1/2$, and $1/8 \le \log 2$. Hence $1+\chi^2(\overline{Q}\|Q_0) \le 2$, so $\chi^2(\overline{Q}\|Q_0) \le 1$. The chi-square to total-variation comparison gives

$$\mathrm{TV}(Q_0,\overline{Q}) \le \frac{1}{2}.$$

Fix any measurable estimator. The two-point testing inequality for the targets 0 under Q_0 and $\Delta/2$ under $\overline{Q}$, with separation threshold $\Delta/4$, gives a tail probability at least $1/4$ under one of these two laws. If the mixture tail is the larger one, one sign ξ has at least that tail probability under $P_\xi^{\otimes n}$. Multiplying the resulting tail probability by $(\Delta/4)^2$ shows that, for this estimator, either P_0 or some P_ξ has mean-squared error at least

$$\left(\frac{\Delta}{4}\right)^2\frac{1}{4} = \frac{\Delta^2}{64}.$$

Taking the supremum over the binary exactly homogeneous class and then the infimum over measurable estimators gives

$$\frac{\Delta^2}{64} \le \inf_{\widehat{\psi}} \sup_{P \in \mathcal{H}^{\mathrm{bin}}_{n,d}(\epsilon)} E_{P^{\otimes n}}\left[\{\widehat{\psi} - \psi(P)\}^2\right].$$

Finally $\Delta^2 = u_\epsilon^2 d/n^2$, so the asserted lower bound follows. $\square$

Lemma 3 records the collision-scale binary statement associated with Zeng et al. [2024]. The overlap-dependent constant u_ϵ is fixed once ϵ is fixed, and the displayed inequality gives a lower bound for estimating $\psi(P)$ over uniform-mass exactly homogeneous binary laws when $d \le n^2$.

A second exact-homogeneity source lemma records the same binary phenomenon with the parametric term included. This statement is used when the transfer argument carries the homogeneous source risk into the real-outcome class.

Lemma 4. [`lem:zeng-binary-exact-homogeneity-lower`] (Exact binary lower bound) *Fix $0 < \epsilon < 1/2$. For integers n, d with $d > 0$, let $\mathcal{H}^{\mathrm{bin}}_{n,d}(\epsilon)$ be the fixed-sample binary-outcome experiment with cell masses $p_k = 1/d$, overlap $\epsilon \le \pi_k \le 1 - \epsilon$, conditional response means in $[0, 1]$, and a common cell treatment effect, and write*

$$\psi(P) = \frac{1}{d} \sum_{k=1}^{d} (\mu_{1k} - \mu_{0k}).$$

There exist constants $a_\epsilon, b_\epsilon > 0$ and an integer N_ϵ such that, for every $n, d \in \mathbb{N}$ with $d > 0$, $N_\epsilon \le n$, and $d \le b_\epsilon n^2$,

$$a_\epsilon \left(\frac{1}{n} + \frac{d}{n^2}\right) \le \inf_{\widehat{\psi}} \sup_{P^{\otimes n} \in \mathcal{H}^{\mathrm{bin}}_{n,d}(\epsilon)} E_P\left[(\widehat{\psi} - \psi(P))^2\right].$$

Proof of Lemma 4. Let the overlap-dependent collision constant be

$$u_\epsilon := \min\left\{\frac{1}{2}, 8\left(\frac{1}{2} - \epsilon\right)^2\right\}, \qquad a_\epsilon := \min\left\{\frac{1}{200}, \frac{u_\epsilon^2}{128}\right\}, \qquad b_\epsilon := 1, \qquad N_\epsilon := 1.$$

Because $0 < \epsilon < 1/2$, we have $u_\epsilon > 0$, and therefore $a_\epsilon > 0$; also $b_\epsilon > 0$.

Fix $n, d \in \mathbb{N}$ with $d > 0$, $N_\epsilon \le n$, and $d \le b_\epsilon n^2$. Then $n > 0$ and, since $b_\epsilon = 1$,

$$d \le n^2.$$

Write

$$R^{\mathrm{bin}}_{n,d,\epsilon} := \inf_{\widehat{\psi}} \sup_{P \in \mathcal{H}^{\mathrm{bin}}_{n,d}(\epsilon)} \mathbb{E}_{P^{\otimes n}}\left[(\widehat{\psi} - \psi(P))^2\right]$$

for the fixed-sample binary exact-homogeneity minimax risk.

First, set

$$\Delta := \frac{2}{5\sqrt{n}}, \qquad \Delta > 0, \qquad \Delta^2 = \frac{4}{25n}.$$

Since $n \ge 1$, $\Delta \le 2/5$, so $1/2 + \Delta \le 1$. Define two endpoint binary laws P_0 and P_1 by uniform cell masses $p_k = 1/d$ and propensities $\pi_k = 1/2$ in every cell, with conditional response means

$$\mu_{0k}(P_0) = \mu_{1k}(P_0) = \frac{1}{2}, \qquad \mu_{0k}(P_1) = \frac{1}{2}, \qquad \mu_{1k}(P_1) = \frac{1}{2} + \Delta$$

for every cell k. The inequalities $0 < \epsilon < 1/2$ put $\pi_k = 1/2$ in $[\epsilon, 1-\epsilon]$, and the preceding bound on Δ puts all displayed conditional means in $[0,1]$. Thus both laws have uniform cell masses, overlap at level ϵ, and a cell effect constant across cells. Hence

$$P_0, P_1 \in \mathcal{H}^{\mathrm{bin}}_{n,d}(\epsilon), \qquad \psi(P_0) = 0, \qquad \psi(P_1) = \Delta.$$

For this endpoint pair, the Bernoulli χ^2 calculation gives total-variation distance at most $1/2$ for the two n-fold product laws because

$$n \frac{(1/2)\Delta^2}{(1/2)(1-1/2)} = \frac{8}{25} \le \log 2.$$

Le Cam's two-point inequality therefore gives

$$\frac{\Delta^2}{16} \le R^{\mathrm{bin}}_{n,d,\epsilon}, \qquad \text{and hence} \qquad \frac{1}{100n} \le R^{\mathrm{bin}}_{n,d,\epsilon}.$$

Second, Lemma 3, applied with $n \ge 1$, $d \ge 1$, $0 < \epsilon < 1/2$, and $d \le n^2$, gives

$$\frac{u_\epsilon^2}{64} \frac{d}{n^2} \le R^{\mathrm{bin}}_{n,d,\epsilon}.$$

If $d \le n$, then

$$\frac{d}{n^2} \le \frac{1}{n}, \qquad \frac{1}{n} + \frac{d}{n^2} \le \frac{2}{n}.$$

Using $a_\epsilon \le 1/200$, we get

$$a_\epsilon \left(\frac{1}{n} + \frac{d}{n^2} \right) \le \frac{1}{200} \frac{2}{n} = \frac{1}{100n} \le R^{\mathrm{bin}}_{n,d,\epsilon}.$$

Otherwise $n \le d$, and hence

$$\frac{1}{n} \le \frac{d}{n^2}, \qquad \frac{1}{n} + \frac{d}{n^2} \le 2\frac{d}{n^2}.$$

Using $a_\epsilon \le u_\epsilon^2/128$, we get

$$a_\epsilon \left(\frac{1}{n} + \frac{d}{n^2} \right) \le \frac{u_\epsilon^2}{128} 2 \frac{d}{n^2} = \frac{u_\epsilon^2}{64} \frac{d}{n^2} \le R^{\mathrm{bin}}_{n,d,\epsilon}.$$

The two cases cover all $d > 0$, and the stated lower bound follows. □

Lemma 4 records the binary exact-homogeneity lower scale associated with Zeng et al. [2024]. The constants depend only on the fixed overlap level, and the active range $d \le b_\epsilon n^2$ is the range transported and then saturated in the all-alphabet real-outcome statement.

The radius-channel converse uses a different binary source: a one-arm problem with arbitrary cell masses and zero control-arm response mean. Its minimax lower bound carries the large-alphabet polynomial scale that appears after the channel attenuation.

Lemma 5. [`lem:zeng-binary-one-arm-lower`] (Binary one-arm lower bound) *For every real overlap constant ϵ, if $0 < \epsilon < 1/2$, then there exist constants $a_\epsilon, b_\epsilon > 0$ and an integer N_ϵ such that the following holds. For every sample size n and alphabet size d,*

- ***(Positive alphabet.)*** *$d \geq 1$.*
- ***(Large sample.)*** *$n \geq N_\epsilon$.*
- ***(Source range.)*** *$d \leq b_\epsilon n \log n$.*

In the fixed-sample binary observed-data experiment with observations $(X, A, Y) \in \{1, \ldots, d\} \times \{0, 1\} \times \{0, 1\}$, arbitrary cell masses $p_k = P(X = k)$, product sampling law $P^{\otimes n}$, overlap

$$\epsilon \leq P(A = 1 \mid X = k) \leq 1 - \epsilon$$

on every positive-mass category, and identically zero control-arm response mean $E_P[Y \mid A = 0, X = k] = 0$, the treated-arm minimax risk satisfies

$$\inf_{\widehat{\psi}} \sup_{P} E_{P^{\otimes n}} \left[\left\{ \widehat{\psi} - \sum_{k=1}^{d} p_k E_P[Y \mid A = 1, X = k] \right\}^2 \right] \geq a_\epsilon \left\{ \frac{1}{n} + \frac{d^2}{n^2 (\log n)^2} \right\},$$

where the infimum is over measurable estimators of the fixed sample.

Proof of Lemma 5. Fix an overlap constant $\epsilon \in (0, 1/2)$. For sample size n and alphabet size d, write $\mathcal{R}_{n,d}(\epsilon)$ for the one-arm minimax risk above, namely

$$\mathcal{R}_{n,d}(\epsilon) = \inf_{\widehat{\psi}} \sup_{P} E_{P^{\otimes n}} \left[\left\{ \widehat{\psi} - \sum_{k=1}^{d} p_k E_P[Y \mid A = 1, X = k] \right\}^2 \right],$$

where the infimum is over measurable fixed-sample estimators and the supremum ranges over the binary control-zero laws described above, with propensities in $[\epsilon, 1 - \epsilon]$. Also write

$$r_{n,d} := \frac{1}{n} + \frac{d^2}{n^2 (\log n)^2}.$$

Set

$$\kappa := \min \left\{ 1, \frac{1 - 2\epsilon}{2\epsilon} \right\}.$$

The overlap algebra gives

$$0 < \kappa \leq 1, \qquad \epsilon(1 + \kappa) \leq 1 - \epsilon.$$

Let

$$A_\epsilon := \frac{\kappa^{14}}{4 \cdot 25600\, (153600000000000)^2}, \qquad K_\epsilon := \frac{10^{50}}{\kappa^{14}}.$$

Both constants are positive. Choose $N_{0,\epsilon}$ so large that, for every $n \geq N_{0,\epsilon}$,

$$n \geq 400, \qquad (\log n)^4 \leq \frac{\kappa^{14}}{10^{50}}\, n.$$

This is the elementary eventual domination of $(\log n)^4$ by n.

We first prove the high-dimensional branch. Suppose that $q \geq 1$, $n \geq N_{0,\epsilon}$, and

$$K_\epsilon (\log n)^4 \leq q \leq n \log n.$$

Because $K_\epsilon \geq 1$ and $n \geq 400$, this implies $q \geq 2$. Put $m = q-1$, and define

$$D_n := \left\lceil \frac{8\log n}{\log 2} \right\rceil, \qquad s_n := \frac{D_n}{128n}, \qquad \rho_{\kappa,n} := \frac{\kappa^5}{10^{10} D_n^2}, \qquad \gamma_\kappa := \frac{1}{10}\frac{2\kappa^2}{(1+\kappa)(2+\kappa)(3+\kappa)}.$$

Thus $D_n \geq 1$, $D_n \leq 20\log n$, and

$$D_n \geq \frac{8\log n}{\log 2}, \qquad \rho_{\kappa,n} > 0, \qquad \frac{\kappa^2}{120} \leq \gamma_\kappa \leq 1.$$

Set

$$\Delta := m\, s_n \rho_{\kappa,n}\gamma_\kappa.$$

The scalar hypotheses needed for the shifted construction are as follows:

$$m s_n \rho_{\kappa,n} \leq 1, \qquad 2m/n^4 \leq 1/8, \qquad 2\frac{m s_n^2}{(\Delta/16)^2} \leq 1/32, \qquad (m s_n)^2 n^{-2} \leq \Delta^2/32.$$

The first bound follows from $m \leq q \leq n\log n$, $\kappa^5 \leq 1$, and $D_n \geq 8\log n/\log 2$. The second follows from $m \leq n\log n \leq n^2$ and $n > 0$, with $n \geq 400$ supplying the remaining numerical slack. For the third, $m \geq q/2 \geq K_\epsilon(\log n)^4/2$,

$$\rho_{\kappa,n}\gamma_\kappa \geq \frac{\kappa^7}{1200000000000\, D_n^2}, \qquad D_n^4 \leq 20^4(\log n)^4,$$

give

$$m\rho_{\kappa,n}^2\gamma_\kappa^2 \geq 2\cdot 16^2\cdot 32,$$

and therefore, using $\Delta = m s_n \rho_{\kappa,n}\gamma_\kappa$,

$$2\frac{m s_n^2}{(\Delta/16)^2} = \frac{2\cdot 16^2}{m\rho_{\kappa,n}^2\gamma_\kappa^2} \leq \frac{1}{32}.$$

For the fourth, the same lower bound on $\rho_{\kappa,n}\gamma_\kappa$, the upper bound on D_n, and $(\log n)^4 \leq \kappa^{14} n/10^{50}$ give

$$32 \leq n^2\rho_{\kappa,n}^2\gamma_\kappa^2,$$

which is equivalent to $(m s_n)^2 n^{-2} \leq \Delta^2/32$. These real inequalities are exactly the nonnegative extended-real budget used for the two discarded good-event complements.

We now construct the finite mixtures used in this branch. Let $a_{\kappa,n} := \kappa^4/(10^{10}D_n^2)$, so $\rho_{\kappa,n} = \kappa a_{\kappa,n}$. The shifted grid consists of the three points $a_{\kappa,n}, 2a_{\kappa,n}, 3a_{\kappa,n}$ and the $2D_n+1$ affine Chebyshev mesh points on $[a_{\kappa,n}, 1]$. Write these positive grid nodes as x_u. For a grid symbol u, augmented by a null symbol $\varnothing$, define

$$p(u) = \begin{cases} 0, & u = \varnothing, \\ s_n x_u, & u \neq \varnothing, \end{cases} \qquad \pi(u) = \begin{cases} \epsilon, & u = \varnothing, \\ \epsilon\left(1 + \rho_{\kappa,n}/x_u\right), & u \neq \varnothing, \end{cases}$$

and

$$\mu(u) = \begin{cases} 1, & u = \varnothing, \\ x_u/(x_u + \rho_{\kappa,n}), & u \neq \varnothing, \end{cases} \qquad f(u) := p(u)\mu(u).$$

The preceding overlap choice gives $\pi(u) \in [\epsilon, 1-\epsilon]$, and the grid construction gives $0 \leq \mu(u) \leq 1$.

Choose signed weights w on the positive grid with total mass zero and positive Jordan mass. Let ω_0 and ω_1 be the normalized positive and negative Jordan parts of w. They have the same moments against the shifted reciprocal-monomial basis, and their shifted target expectations satisfy

$$\gamma_\kappa \le \sum_u \omega_0(u)\frac{x_u}{x_u+\rho_{\kappa,n}} - \sum_u \omega_1(u)\frac{x_u}{x_u+\rho_{\kappa,n}}.$$

Now inverse-tilt these two laws by the factor $\rho_{\kappa,n}/x_u$ and put the remaining mass at $\varnothing$:

$$\nu_t(\{u\}) = \omega_t(u)\frac{\rho_{\kappa,n}}{x_u}, \qquad \nu_t(\{\varnothing\}) = 1 - \sum_u \omega_t(u)\frac{\rho_{\kappa,n}}{x_u}, \qquad t=0,1.$$

Because $\rho_{\kappa,n} \le x_u$ on the shifted grid, these are probability laws. For every $i+j+k \le D_n$, the test

$$x \longmapsto x^{i+k}(x+\rho_{\kappa,n})^j/x$$

is represented by that basis with degree at most D_n. The Jordan moment identities therefore transport through the inverse tilt to

$$\int p(u)^i\{p(u)\pi(u)\}^j\{p(u)\pi(u)\mu(u)\}^k\,d\nu_0(u) = \int p(u)^i\{p(u)\pi(u)\}^j\{p(u)\pi(u)\mu(u)\}^k\,d\nu_1(u).$$

The same inverse-tilt identities give common active mass and separated treated-functional atoms,

$$\int p\,d\nu_0 = \int p\,d\nu_1 = s_n\rho_{\kappa,n}, \qquad s_n\rho_{\kappa,n}\gamma_\kappa \le \int f\,d\nu_0 - \int f\,d\nu_1.$$

Let $z=(z_1,\dots,z_m)$ have product law $\nu_t^{\otimes m}$ under side $t\in\{0,1\}$, and set

$$\theta_t := m\int f\,d\nu_t, \qquad \alpha := 1 - ms_n\rho_{\kappa,n}, \qquad \delta := \Delta/16.$$

By the first scalar budget, $\alpha \ge 0$. Also $\Delta > 0$ by positivity of its factors, and the active-mass cap $ms_n\rho_{\kappa,n} \le 1$, together with $\gamma_\kappa \le 1$, gives $\Delta \le 1$. Hence

$$0<\delta<1, \qquad \delta \le \frac{1}{2}.$$

Define the relaxed good event

$$G_t := \left\{ z : 0 < \alpha + \sum_{i=1}^m p(z_i), \quad \left|\alpha + \sum_{i=1}^m p(z_i) - 1\right| \le \delta, \quad \left|\sum_{i=1}^m f(z_i) - \theta_t\right| \le \delta \right\}.$$

The two-statistic Chebyshev envelope, applied with the displayed $0<\delta<1$, and the displayed budget give

$$\nu_t^{\otimes m}(G_t^c) \le 2\frac{ms_n^2}{\delta^2} \le \frac{1}{32}, \qquad t=0,1,$$

so each G_t has positive probability.

For $z \in G_t$, form a binary control-zero law $P_{t,z}$ on $q=m+1$ cells by assigning unnormalized cell masses $\alpha, p(z_1),\dots,p(z_m)$, anchor propensity ϵ, active propensities $\pi(z_i)$, anchor treated mean zero, active treated means $\mu(z_i)$, and then normalizing by $S_z := \alpha + \sum_i p(z_i)$. The event

G_t gives $S_z > 0$, the preceding bounds give the required overlap and binary means, and all control-arm means are zero. Moreover

$$|\psi(P_{t,z}) - \theta_t| \leq |S_z - 1| + \left| \sum_{i=1}^{m} f(z_i) - \theta_t \right| \leq 2\delta = \Delta/8.$$

The separation display for ν_0, ν_1 gives

$$\Delta \leq |\theta_0 - \theta_1|.$$

For each realization z, let K_z be the product law of independent triple-counts over the anchor and active cells with Poisson rates

$$(0, 4n\alpha\epsilon, 4n\alpha(1-\epsilon))$$

on the anchor cell and

$$(4np(z_i)\pi(z_i)\mu(z_i),\ 4np(z_i)\pi(z_i)(1-\mu(z_i)),\ 4np(z_i)(1-\pi(z_i)))$$

on active cell i. The logarithmic degree supplies, for every grid atom,

$$4\{4np(u)\} + 2\log(n^2) \leq (D_n + 1)\log 2.$$

Together with the raw moment identities through degree D_n, this Taylor budget bounds the one-coordinate mixed triple-Poisson total variation by $2/n^4$. Tensorization over the m active coordinates, with the common anchor law factored out, gives the unconditioned predictive bound

$$\mathrm{TV}\left(\int K_z \, d\nu_0^{\otimes m}(z), \int K_z \, d\nu_1^{\otimes m}(z) \right) \leq 2m/n^4 \leq 1/8.$$

After restricting $\nu_t^{\otimes m}$ to G_t and renormalizing, the loss in total variation is bounded by the two discarded probabilities. Hence the conditioned predictive mixtures satisfy

$$\mathrm{TV}\left(\int_{G_0} K_z \, d\nu_0^{\otimes m}(z \mid G_0), \int_{G_1} K_z \, d\nu_1^{\otimes m}(z \mid G_1) \right) \leq \frac{1}{4}.$$

For every $z \in G_t$, the Poissonized histogram generated from $P_{t,z}$ at total intensity $4nS_z$, after recording the three arm-outcome labels in each cell, maps to exactly the kernel K_z. Also $S_z \geq 1/2$, because $\delta \leq 1/2$. Therefore the total Poisson count $N_z \sim \mathrm{Pois}(4nS_z)$ satisfies $4nS_z \geq 2n$, and the Poisson lower-tail bound gives $\Pr\{N_z < n\} \leq n^{-2}$. Since $|\theta_t| \leq ms_n$, the last scalar budget yields

$$\theta_t^2 \Pr\{N_z < n\} \leq (ms_n)^2 n^{-2} \leq \Delta^2/32, \qquad (z \in G_t,\ t \in \{0, 1\}).$$

We use the following finite-mixture testing reduction. If two finite mixtures have side centers θ_0, θ_1, all component targets lie within radius r of their side center, $2s \leq |\theta_0 - \theta_1|$, their predictive total variation is at most c, and the Poisson truncation penalty on each side is bounded by η, then

$$\frac{s^2\{(1-c)/2\} - 2r^2 - \eta}{2} \leq \mathcal{R}_{n,q}(\epsilon).$$

Apply this with

$$s = \Delta/2, \qquad r = \Delta/8, \qquad c = 1/4, \qquad \eta = \Delta^2/32.$$

The numerical identity

$$\frac{(\Delta/2)^2\{(1-1/4)/2\} - 2(\Delta/8)^2 - \Delta^2/32}{2} = \frac{\Delta^2}{64}$$

gives

$$\frac{\Delta^2}{64} \leq \mathcal{R}_{n,q}(\epsilon).$$

The calibrated signal inequality

$$\frac{\kappa^7}{153600000000000\, nD_n} \leq s_n \rho_{\kappa,n} \gamma_\kappa$$

and $D_n \leq 20 \log n$ imply

$$\frac{\kappa^{14}}{25600(153600000000000)^2} \frac{m^2}{n^2(\log n)^2} \leq \mathcal{R}_{n,q}(\epsilon).$$

Since $m \geq q/2$, this proves

$$A_\epsilon \frac{q^2}{n^2(\log n)^2} \leq \mathcal{R}_{n,q}(\epsilon)$$

throughout the range $K_\epsilon (\log n)^4 \leq q \leq n \log n$.

The parametric component supplies the fixed-sample term. Fix $n \geq 1$ and $d \geq 1$, and put

$$\eta_n := \frac{2/5}{\sqrt{n}}.$$

Since $\eta_n^2 = 4/(25n)$ and $\eta_n \leq 2/5$, the two endpoint binary laws with propensity $1/2$, treated means $1/2$ and $1/2 + \eta_n$, and baseline control mean $1/2$ are valid. Deterministically replacing every control-arm outcome by zero preserves the treated functional and contracts total variation; it yields two control-zero laws Q_0, Q_1 with overlap in $[\epsilon, 1-\epsilon]$, treated functionals

$$\psi(Q_0) = \frac{1}{2}, \qquad \psi(Q_1) = \frac{1}{2} + \eta_n,$$

and product total variation at most $1/2$. The total-variation bound is calibrated by

$$n \frac{(1/2)\eta_n^2}{(1/2)(1-1/2)} = \frac{8}{25} \leq \log 2.$$

Applying the two-point testing inequality with half-separation $\eta_n/2$ and total variation $1/2$ gives

$$\frac{1}{100n} = \left(\frac{\eta_n}{2}\right)^2 \frac{1-1/2}{2} \leq \mathcal{R}_{n,d}(\epsilon).$$

Now choose N_ϵ so large that $n \geq N_\epsilon$ implies $n \geq N_{0,\epsilon}$, $n \geq 3$, and

$$(\log n)^6 \leq \frac{n}{K_\epsilon^2 + 1}.$$

Set

$$a_\epsilon := \min\{1/200, A_\epsilon/2\}, \qquad b_\epsilon := 1.$$

Then $a_\epsilon, b_\epsilon > 0$. Fix $n \ge N_\epsilon$, $d \ge 1$, and $d \le b_\epsilon n \log n = n \log n$. If

$$K_\epsilon (\log n)^4 \le d,$$

the high-dimensional bound and the parametric bound give

$$a_\epsilon \frac{d^2}{n^2 (\log n)^2} \le \frac{1}{2} \mathcal{R}_{n,d}(\epsilon), \qquad \frac{a_\epsilon}{n} \le \frac{1}{2} \mathcal{R}_{n,d}(\epsilon),$$

and hence $a_\epsilon r_{n,d} \le \mathcal{R}_{n,d}(\epsilon)$. In the complementary branch, $d < K_\epsilon (\log n)^4$. The choice of N_ϵ gives

$$K_\epsilon^2 (\log n)^6 \le n,$$

so

$$\frac{d^2}{n^2 (\log n)^2} \le \frac{1}{n}, \qquad r_{n,d} \le \frac{2}{n}.$$

Since $a_\epsilon \le 1/200$, the parametric lower bound yields

$$a_\epsilon r_{n,d} \le \frac{1}{100n} \le \mathcal{R}_{n,d}(\epsilon).$$

Thus for all $n \ge N_\epsilon$, $d \ge 1$, and $d \le b_\epsilon n \log n$,

$$a_\epsilon \left\{ \frac{1}{n} + \frac{d^2}{n^2 (\log n)^2} \right\} \le \mathcal{R}_{n,d}(\epsilon),$$

which is the asserted one-arm lower bound. □

Lemma 5 formalizes the arbitrary-mass binary rare-cell lower scale associated with Zeng et al. [2024]. In the main lower-bound construction, the hypothesis-independent channel in Algorithm 4 attenuates this binary separation by the radius factor before the affine scaling by M.

The final source lemma is an occupancy reciprocal bound. It controls the event that no empirically crossed cell appears and the reciprocal of the crossed-cell count used by the collision estimator.

Lemma 6. [`lem:zeng-usable-occupancy-reciprocal`] (Usable occupancy reciprocal) *For every overlap parameter $\epsilon \in \mathbb{R}$, there are constants $b_\epsilon, B_\epsilon > 0$, depending only on ϵ, such that the following holds. For every $n, d \in \mathbb{N}$, every $M, \sigma \in \mathbb{R}$, and every $P \in \mathcal{P}_{d,\epsilon,M,\sigma}$, let $Q_P^{(n)}$ denote the n-fold observed-data law and define*

$$N_{ak} := \sum_{i=1}^{n} \mathbf{1}\{X_i = k, A_i = a\}, \qquad N_k := N_{0k} + N_{1k}, \qquad I_k := \mathbf{1}\{N_{0k} > 0,\ N_{1k} > 0\},$$

with

$$R_n := \sum_{k=1}^{d} N_k I_k.$$

Then

$$Q_P^{(n)}(R_n = 0) \le 2\exp\left\{-b_\epsilon \frac{n^2}{n \vee d}\right\},$$

and

$$E_{Q_P^{(n)}}\left[\mathbf{1}\{R_n > 0\}\frac{1}{R_n}\right] \le B_\epsilon \left\{\frac{n \vee d}{n^2} + \exp\left(-b_\epsilon \frac{n^2}{n \vee d}\right)\right\}.$$

Proof of Lemma 6. On the active range $0 < \epsilon < 1/2$, let $c_\epsilon > 0$ be the constant in the independent-Poisson usable-occupancy Laplace bound, and set

$$b_\epsilon := c_\epsilon/8, \qquad B_\epsilon := \max\{64/c_\epsilon, 2\}.$$

Set $b_\epsilon = B_\epsilon = 1$ on the complementary range. For any $P \in \mathcal{P}_{d,\epsilon,M,\sigma}$, Definition 1 gives $0 < \epsilon < 1/2$, so the active values of the constants apply.

If $d = 0$, the cell-mass identity gives $0 = 1$, so the assertion is immediate. Suppose $n = 0$ and hence, after the preceding case, $d > 0$. The zero-event mass is at most one and therefore at most $2\exp\{0\}$. The displayed rate is interpreted at this endpoint with the real-division convention $x/0 = 0$, so $(n \vee d)/n^2 = 0$ and $\exp\{-b_\epsilon n^2/(n \vee d)\} = 1$. The reciprocal integrand $\mathbf{1}\{R_0 > 0\}/R_0$ is identically zero, while the right-hand side is $B_\epsilon > 0$.

Assume now that $n, d > 0$. Put

$$\kappa_{n,d} := c_\epsilon \frac{(n/2)^2}{\max\{n/2, d\}}, \qquad s_{n,d} := \frac{n^2}{n \vee d}.$$

Since $\max\{n/2, d\} \le \max\{n, d\}$,

$$\frac{c_\epsilon}{4} s_{n,d} \le \kappa_{n,d}.$$

Let Q be the law of the observed cell-arm mark (X, A). In the half-intensity marked-Poisson experiment with mark law Q and total intensity $n/2$, let $\widetilde{R}$ be the usable total obtained after regrouping the independent arm-cell counts. If $p_k = P(X = k)$ and $x_k = (n/2)p_k$, then the two arm-cell intensities in cell k are $(n/2)Q\{(k,0)\}$ and $(n/2)Q\{(k,1)\}$. They sum to x_k, the total cell intensities satisfy $\sum_k x_k = n/2$, and fixed overlap gives

$$\epsilon x_k \le (n/2)Q\{(k,0)\}, \qquad \epsilon x_k \le (n/2)Q\{(k,1)\}$$

for every positive-mass cell, with zero-mass cells contributing trivially. The independent-Poisson Laplace bound for these intensities gives

$$E\exp\{-\widetilde{R}\} \le \exp\left\{-c_\epsilon \frac{(n/2)^2}{\max\{n/2, d\}}\right\} = \exp\{-\kappa_{n,d}\}.$$

For each integer $r \ge 0$,

$$\mathbf{1}\{r = 0\} \le e^{-r}, \qquad \begin{cases} 1, & r = 0, \\ 1/r, & r > 0 \end{cases} \le \frac{2}{\kappa_{n,d}} + e^{\kappa_{n,d}/2}e^{-r},$$

where the second inequality splits the positive case according to whether $r \ge \kappa_{n,d}/2$. Integrating these deterministic bounds and using the Laplace display yields

$$\Pr(\widetilde{R} = 0) \le e^{-\kappa_{n,d}}, \qquad E\left[\begin{cases} 1, & \widetilde{R} = 0, \\ 1/\widetilde{R}, & \widetilde{R} > 0 \end{cases}\right] \le \frac{2}{\kappa_{n,d}} + e^{-\kappa_{n,d}/2}.$$

Let R_m^{str} denote the usable total in the first m marks of an i.i.d. stream with one-step law Q. Adding observations can only increase the usable total, so both $\mathbf{1}\{R_m^{\mathrm{str}} = 0\}$ and the zero-plus-reciprocal penalty

$$\begin{cases} 1, & R_m^{\mathrm{str}} = 0, \\ 1/R_m^{\mathrm{str}}, & R_m^{\mathrm{str}} > 0 \end{cases}$$

are nonincreasing in m. If $N \sim \mathrm{Poisson}(n/2)$, then Markov's inequality gives

$$\Pr(N \geq n+1) \leq \frac{n/2}{n+1} \leq \frac{1}{2}, \qquad \text{hence} \qquad \Pr(N \leq n) \geq \frac{1}{2}.$$

For either of the two nonnegative monotone prefix functionals g_m above,

$$\Pr(N \leq n) E g_n \leq E g_N,$$

because $g_n \leq g_m$ on the event $m \leq n$. Therefore $E g_n \leq 2 E g_N$. The random-prefix expectation $E g_N$ is exactly the corresponding marked-Poisson expectation with total intensity $n/2$, and the deterministic auxiliary outcome mark does not change the prefix statistic. The fixed-prefix bounds are thus

$$E\mathbf{1}\{R_n^{\mathrm{str}} = 0\} \leq 2e^{-\kappa_{n,d}}, \qquad E\left[\begin{cases} 1, & R_n^{\mathrm{str}} = 0, \\ 1/R_n^{\mathrm{str}}, & R_n^{\mathrm{str}} > 0 \end{cases}\right] \leq 2\left\{\frac{2}{\kappa_{n,d}} + e^{-\kappa_{n,d}/2}\right\}.$$

The n-prefix law of the cell-arm stream is the observed mark marginal of $Q_P^{(n)}$, and forgetting outcomes identifies the fixed-sample usable total with the corresponding mark-prefix usable total. Hence

$$Q_P^{(n)}(R_n = 0) \leq 2e^{-\kappa_{n,d}}$$

and, because $\mathbf{1}\{R_n > 0\}/R_n$ is bounded by the zero-plus-reciprocal penalty,

$$E_{Q_P^{(n)}}\left[\mathbf{1}\{R_n > 0\}\frac{1}{R_n}\right] \leq 2\left\{\frac{2}{\kappa_{n,d}} + e^{-\kappa_{n,d}/2}\right\}.$$

Using $\kappa_{n,d} \geq (c_\epsilon/4)s_{n,d}$ and $b_\epsilon = c_\epsilon/8$,

$$Q_P^{(n)}(R_n = 0) \leq 2\exp\{-b_\epsilon s_{n,d}\}.$$

For the reciprocal term,

$$2\left\{\frac{2}{\kappa_{n,d}} + e^{-\kappa_{n,d}/2}\right\} \leq \frac{64}{c_\epsilon}\frac{n \vee d}{n^2} + 2\exp\{-b_\epsilon s_{n,d}\} \leq B_\epsilon\left\{\frac{n \vee d}{n^2} + \exp(-b_\epsilon s_{n,d})\right\},$$

by the choice $B_\epsilon \geq \max\{64/c_\epsilon, 2\}$. This is the asserted reciprocal inequality. □

Lemma 6 records the occupancy statement associated with the collision calculations of Zeng et al. [2024]. The two displayed inequalities bound, respectively, the zero-crossing probability and the expected reciprocal occupancy on the crossed-cell event. These controls are the probabilistic ingredients behind the collision estimator's dependence on d/n^2 and fixed overlap.

B Proofs for upper bounds

This appendix records the auxiliary inequalities that support the constructive upper bounds in Theorems 1 and 2. The proof has two mechanisms. The polynomial branch first controls the shifted-Chebyshev coefficients, then bounds same-cell and distinct-cell factorial covariances, then combines those bounds with a deterministic heavy-set plug-in argument. The collision branch controls the occupancy-weighted treated-control contrast on empirically crossed cells. These mechanisms correspond to the two estimator branches introduced in Algorithms 1 and 2.

The first ingredient is a coefficient envelope for the shifted-Chebyshev polynomial used in the light-cell component of Algorithm 1. It converts the polynomial coefficients into a scale bound on cell masses below the light-cell threshold, matching the large-alphabet approximation strategy of Jiao et al. [2015], Wu and Yang [2016].

Lemma 7. [`lem:shifted-chebyshev-coefficient-envelope`] (Shifted coefficient envelope) *Let K be a positive integer, let $B, p \in \mathbb{R}$, and define, for $0 \le j \le K-2$,*

$$g_{Kj} := (-1)^j \frac{2^{2j+3}}{K(K+j+2)} \binom{K+j+2}{2j+4}.$$

Assume:

- ***(Positive scale.)*** *$B > 0$.*
- ***(Nonnegative mass.)*** *$p \ge 0$.*
- ***(Budget bound.)*** *$p \le B$.*

Then

$$\sum_{j=0}^{K-2} \left| \frac{g_{Kj}}{B^{j+1}} \right| p^{j+2} \le B\, 6^K.$$

Proof of Lemma 7. Define the absolute coefficient envelope

$$\mathcal{G}_K(x) := \sum_{j=0}^{K-2} |g_{Kj}| x^j.$$

Let T_K be the Chebyshev polynomial of the first kind. The shifted expansion gives

$$T_K(1-2x) = 1 - 2K^2 x + 2K^2 x^2 \sum_{j=0}^{K-2} g_{Kj} x^j.$$

This follows by separating the linear term in the standard shifted-Chebyshev expansion

$$T_K(1-2x) = 1 + \sum_{\ell=1}^{K} (-1)^\ell \frac{K}{K+\ell} \binom{K+\ell}{2\ell} 4^\ell x^\ell$$

and reindexing the terms $\ell = j+2$, whose coefficients are $2K^2 g_{Kj}$.

Since the sign of g_{Kj} is $(-1)^j$, one has

$$|g_{Kj}| = (-1)^j g_{Kj}, \qquad \mathcal{G}_K(z) = \sum_{j=0}^{K-2} g_{Kj} (-z)^j.$$

Putting $x = -1$ in the shifted identity therefore yields

$$T_K(3) = 1 + 2K^2 + 2K^2\mathcal{G}_K(1).$$

Because $\mathcal{G}_K(1) \geq 0$ and $K \geq 1$,

$$\mathcal{G}_K(1) \leq T_K(3).$$

The Chebyshev recurrence $T_{n+2}(3) = 6T_{n+1}(3) - T_n(3)$, with the initial cases $T_0(3) = 1$ and $T_1(3) = 3$, gives

$$T_K(3) \leq 6^K.$$

If $0 \leq x \leq 1$, then termwise monotonicity gives

$$\mathcal{G}_K(x) = \sum_{j=0}^{K-2} |g_{Kj}| x^j \leq \sum_{j=0}^{K-2} |g_{Kj}| = \mathcal{G}_K(1) \leq 6^K.$$

Now set

$$x := \frac{p}{B}.$$

The assumptions $B > 0$, $p \geq 0$, and $p \leq B$ imply

$$0 \leq x \leq 1.$$

For each j, using $B > 0$,

$$\left| \frac{g_{Kj}}{B^{j+1}} \right| p^{j+2} = \frac{p^2}{B} |g_{Kj}| \left(\frac{p}{B} \right)^j.$$

Summing this identity over $j = 0, \ldots, K-2$ gives

$$\sum_{j=0}^{K-2} \left| \frac{g_{Kj}}{B^{j+1}} \right| p^{j+2} = \frac{p^2}{B} \mathcal{G}_K \left(\frac{p}{B} \right).$$

The envelope bound at $x = p/B$ and the inequality $p^2 \leq B^2$ give

$$\sum_{j=0}^{K-2} \left| \frac{g_{Kj}}{B^{j+1}} \right| p^{j+2} \leq \frac{p^2}{B} 6^K \leq B 6^K.$$

□

Lemma 7 gives the deterministic side of the light-cell calculation. Once a cell mass is below the approximation scale B, the entire signed coefficient expansion is controlled by $B6^K$, so the later variance bounds can aggregate the polynomial terms without tracking each coefficient separately.

The next two moment bounds handle the factorial statistics that estimate the light-cell polynomial terms. They are stated for an independent block of size m with product law $Q_P^{(m)}$, the block analogue of the observed product experiment in Definition 3. The same-cell bound includes the extra covariance contribution created when two statistics use the same adjustment cell, and the distinct-cell bound records the simpler interaction across different cells.

Lemma 8. [`lem:same-cell-factorial-cross-moment`] (Same-cell cross moment) *Let $P \in \mathcal{P}_{d,\epsilon,M,\sigma}$ be as in Definition 1, let $k \in \{1,\dots,d\}$, and write $p_k = P(X = k)$ for the cell mass. For a block of m independent observations with joint law $Q_P^{(m)}$, write $V_i = Y_i/M$. For an arm $c \in \{0,1\}$ and an order $q \in \mathbb{N}$, define*

$$U_{k,c,q} := \frac{1}{(m)_{q+2}} \sum_{i_0,\dots,i_{q+1} \text{ distinct}} V_{i_0} \mathbf{1}\{X_{i_0} = k, A_{i_0} = c\} \mathbf{1}\{X_{i_1} = k\} \prod_{t=2}^{q+1} \mathbf{1}\{X_{i_t} = k, A_{i_t} = c\}.$$

Assume:

- ***(First order.)*** *$j + 2 \le K$.*
- ***(Second order.)*** *$r + 2 \le K$.*
- ***(Block size.)*** *$4(K+2)^2 \le m$.*

Then, for all arms $a, b \in \{0,1\}$,

$$\left|\mathrm{Cov}_{Q_P^{(m)}}(U_{k,a,j}, U_{k,b,r})\right| \le \frac{2(K+2)^2}{m} p_k^{j+2} p_k^{r+2} + \exp(1) \left(\frac{5}{4} p_k\right)^{r+2} \left(\frac{5}{4} p_k + \frac{r+2}{m}\right)^{j+2}.$$

Proof of Lemma 8. Write

$$p := p_k = P(X = k).$$

For an observed record $O = (X, A, Y)$, write $V = Y/M$. Define the left coordinate factors

$$L_q(O) := \begin{cases} V\mathbf{1}\{X = k, A = a\}, & q = 0, \\ \mathbf{1}\{X = k\}, & q = 1, \\ \mathbf{1}\{X = k, A = a\}, & 2 \le q \le j+1, \end{cases}$$

and define the right coordinate factors R_s in the same way, with (b, r, s) in place of (a, j, q).

For a partial matching N of size h between $\{0,\dots,j+1\}$ and $\{0,\dots,r+1\}$, let T_N be the merged coordinate set, and let ι_L and ι_R map left and right coordinates into T_N. Thus $\iota_R(s) = \iota_L(q)$ exactly when N matches q with s. Put

$$\gamma_N := \frac{(m)_{(j+2)+(r+2)-h}}{(m)_{j+2}(m)_{r+2}},$$

$$F_{N,t}(O) := \prod_{q:\, \iota_L(q)=t} L_q(O) \prod_{s:\, \iota_R(s)=t} R_s(O), \qquad M_N := \int \prod_{t \in T_N} F_{N,t}(O_t)\, dP_O^{\otimes T_N}((O_t)_{t \in T_N}).$$

Finally set

$$m_{k,a,j} := \int \prod_{q=0}^{j+1} L_q(O_q)\, dP_O^{\otimes(j+2)}((O_q)_q), \qquad m_{k,b,r} := \int \prod_{s=0}^{r+1} R_s(O_s)\, dP_O^{\otimes(r+2)}((O_s)_s).$$

The block-size condition implies $K \le m$, hence $j + 2 \le m$ and $r + 2 \le m$. The selector factors are measurable bounded indicators, and the marked factor is integrable under the moment conditions in Definition 1; the merged kernels are finite products with at most two marked factors at any single observed coordinate, hence integrable. With the coordinate factors just defined, each statistic $U_{k,a,j}$ and $U_{k,b,r}$ is its normalized ordered-product sum over injective maps into the block. Expanding the product of the two normalized sums and grouping each ordered pair of injective maps by the partial matching N that records their shared block indices gives a contribution $\gamma_N M_N$ for matching pattern N. The empty matching has product moment $m_{k,a,j}m_{k,b,r}$ by independence and is then centered by subtracting $m_{k,a,j}m_{k,b,r}$. Therefore

$$\mathrm{Cov}(U_{k,a,j}, U_{k,b,r}) = (\gamma_\varnothing - 1)m_{k,a,j}m_{k,b,r} + \sum_{h=1}^{\min(j+2,r+2)} \sum_{N\in\mathcal{M}_{j,r}(h)} \gamma_N\, M_N,$$

where $\mathcal{M}_{j,r}(h)$ is the finite set of partial matchings of size h.

Each one-coordinate mean is bounded by the cell mass:

$$\left|\int L_q(O)\, dP_O(O)\right| \le p, \qquad \left|\int R_s(O)\, dP_O(O)\right| \le p.$$

For the outcome-marked coordinate, the argument splits according to the cell mass. If $p = 0$, the observed arm-cell event $\{X = k, A = c\}$ has P_O-mass zero for each arm c, so the marked integral is zero. If $p > 0$, the supported-cell branch of Assumption 4 applies to the arm-cell outcome law for each arm c, and Assumption 3 supplies positive observed arm-cell mass when this law is read as the observed conditional law. The mean bound and $M \ge 1$ from Definition 1 give absolute normalized arm-cell mean at most $1/2$. The observed arm-cell restriction has total mass $p\{\pi_k \mathbf{1}_{\{c=1\}} + (1 - \pi_k)\mathbf{1}_{\{c=0\}}\} \le p$, so the absolute marked integral is bounded by $p/2$. The selector-only coordinates integrate to cell or arm-cell probabilities bounded by p. Since the product-law means factor coordinatewise,

$$|m_{k,a,j}| \le p^{j+2}, \qquad |m_{k,b,r}| \le p^{r+2}.$$

The empty-matching normalization satisfies, under $j + 2 \le K$, $r + 2 \le K$, and $4(K + 2)^2 \le m$,

$$|\gamma_\varnothing - 1| \le \frac{2(K+2)^2}{m}.$$

Therefore

$$|(\gamma_\varnothing - 1)m_{k,a,j}m_{k,b,r}| \le \frac{2(K+2)^2}{m} p^{j+2}p^{r+2}.$$

It remains to bound the positive-matching part. For each merged coordinate $t \in T_N$, the fiber definition of $F_{N,t}$ gives exactly one of the following forms: a single left factor L_q, a single right factor R_s, or a matched product L_qR_s. The first two cases include the lone unmarked cell-indicator coordinate. For single factors, the preceding one-coordinate bound gives

$$\left|\int L_q\, dP_O\right| \le p \le \frac{5}{4}p, \qquad \left|\int R_s\, dP_O\right| \le p \le \frac{5}{4}p.$$

For matched products, first note the square-moment bounds

$$\int L_q(O)^2\, dP_O(O) \le \frac{5}{4}p, \qquad \int R_s(O)^2\, dP_O(O) \le \frac{5}{4}p.$$

For the outcome-marked coordinate these bounds again split on p: the zero-mass branch is null on every arm-cell event, while on $p > 0$, Assumption 3 supplies the positive arm-cell support for either arm. The arm-cell law then satisfies the mean and central second-moment bounds from Assumptions 4 and 6; expanding the second moment around μ_{ck} gives an unnormalized second moment at most $5M^2/4$, hence, using $M \geq 1$ from Definition 1, a normalized second moment at most $5/4$. Multiplying by the observed arm-cell mass, which is at most p, gives the displayed square-moment bound for the marked coordinate. Selector coordinates are idempotent indicators with event mass at most p. Hence, by integrability, $|\int fg| \leq \int |fg|$ and $2|fg| \leq f^2 + g^2$,

$$\left| \int L_q(O) R_s(O)\, dP_O(O) \right| \leq \frac{1}{2} \int \{L_q(O)^2 + R_s(O)^2\}\, dP_O(O) \leq \frac{5}{4} p.$$

Consequently every merged coordinate satisfies

$$\left| \int F_{N,t}(O)\, dP_O(O) \right| \leq \frac{5}{4} p.$$

The merged product moment factors over the merged coordinates:

$$M_N = \prod_{t \in T_N} \int F_{N,t}(O)\, dP_O(O).$$

Since $|T_N| = (j+2) + (r+2) - h$,

$$|M_N| \leq \left(\frac{5}{4} p \right)^{(j+2)+(r+2)-h}.$$

Using $\gamma_N \geq 0$ and the triangle inequality,

$$\left| \sum_{h=1}^{\min(j+2,r+2)} \sum_{N \in \mathcal{M}_{j,r}(h)} \gamma_N M_N \right| \leq \sum_{h=1}^{\min(j+2,r+2)} \sum_{N \in \mathcal{M}_{j,r}(h)} \gamma_N \left(\frac{5}{4} p \right)^{(j+2)+(r+2)-h}.$$

It remains only to sum the normalization weights. Let $u = j+2$ and $v = r+2$. For each h, the size-h normalization bound and the partial-matching count give

$$\sum_{N \in \mathcal{M}_{j,r}(h)} \gamma_N \leq \binom{u}{h} v^h \frac{\exp(1)}{m^h}.$$

Here $|\mathcal{M}_{j,r}(h)| = \binom{u}{h}\binom{v}{h} h!$, and the displayed bound uses $\binom{v}{h} \leq v^h/h!$ together with the factorial-normalization estimate supplied by the block-size condition $4(K+2)^2 \leq m$. Therefore, for every $x \geq 0$,

$$\begin{aligned}
&\sum_{h=1}^{\min(u,v)} \sum_{N \in \mathcal{M}_{j,r}(h)} \gamma_N x^{u+v-h} \\
&\quad \leq \sum_{h=1}^{\min(u,v)} \exp(1) x^v \binom{u}{h} x^{u-h} \left(\frac{v}{m} \right)^h \\
&\quad \leq \sum_{h=0}^{u} \exp(1) x^v \binom{u}{h} x^{u-h} \left(\frac{v}{m} \right)^h = \exp(1) x^v \left(x + \frac{v}{m} \right)^u.
\end{aligned}$$

The choice $x = (5/4)p$ is admissible because $p \geq 0$, and it yields

$$\left| \sum_{h=1}^{\min(j+2,r+2)} \sum_{N \in \mathcal{M}_{j,r}(h)} \gamma_N M_N \right| \leq \exp(1) \left(\frac{5}{4} p \right)^{r+2} \left(\frac{5}{4} p + \frac{r+2}{m} \right)^{j+2}.$$

Combining the empty-matching and positive-matching bounds by the triangle inequality gives

$$\left| \mathrm{Cov}_{Q_P^{(m)}}(U_{k,a,j}, U_{k,b,r}) \right| \leq \frac{2(K+2)^2}{m} p_k^{j+2} p_k^{r+2} + \exp(1) \left(\frac{5}{4} p_k \right)^{r+2} \left(\frac{5}{4} p_k + \frac{r+2}{m} \right)^{j+2},$$

as claimed. □

Lemma 9. [`lem:distinct-cell-factorial-cross-moment`] (Distinct-cell cross moment) *Let $P \in \mathcal{P}_{d,\epsilon,M,\sigma}$ be a model in Definition 1, and let $O_1, \ldots, O_m$ be an independent block with joint law $Q_P^{(m)}$. For $k \in \{1, \ldots, d\}$, $a \in \{0, 1\}$, and $j \geq 0$, let $U_{k,a,j}$ be the ordered distinct-index one-mark factorial statistic from Lemma 8, equivalently*

$$U_{k,a,j} = \frac{1}{(m)_{j+2}} \sum_{\substack{i_0, \ldots, i_{j+1} \in \{1, \ldots, m\} \\ \text{all distinct}}} \frac{Y_{i_0}}{M} \mathbf{1}\{X_{i_0} = k, A_{i_0} = a\} \mathbf{1}\{X_{i_1} = k\} \prod_{q=2}^{j+1} \mathbf{1}\{X_{i_q} = k, A_{i_q} = a\}.$$

Assume:

- ***(Distinct cells.)*** *$k \neq \ell$.*
- ***(Admissible orders.)*** *$j + 2 \leq K$ and $r + 2 \leq K$.*
- ***(Block size.)*** *$4(K+2)^2 \leq m$.*

Then, for all arms $a, b \in \{0, 1\}$,

$$\left| \mathrm{Cov}_{Q_P^{(m)}}(U_{k,a,j}, U_{\ell,b,r}) \right| \leq \frac{2(K+2)^2}{m} p_k^{j+2} p_\ell^{r+2}.$$

Proof of Lemma 9. Let P_O be the one-observation marginal of $Q_P^{(m)}$, and for an observed record $O = (X, A, Y)$ write $V = Y/M$. Define the left coordinate factors

$$L_q(O) := \begin{cases} V \mathbf{1}\{X = k, A = a\}, & q = 0, \\ \mathbf{1}\{X = k\}, & q = 1, \\ \mathbf{1}\{X = k, A = a\}, & 2 \leq q \leq j+1, \end{cases}$$

and the right coordinate factors

$$R_s(O) := \begin{cases} V \mathbf{1}\{X = \ell, A = b\}, & s = 0, \\ \mathbf{1}\{X = \ell\}, & s = 1, \\ \mathbf{1}\{X = \ell, A = b\}, & 2 \leq s \leq r+1. \end{cases}$$

Put

$$\mu_{k,a,j} := \int \prod_{q=0}^{j+1} L_q(O_q)\, dP_O^{\otimes(j+2)}((O_q)_q), \qquad \mu_{\ell,b,r} := \int \prod_{s=0}^{r+1} R_s(O_s)\, dP_O^{\otimes(r+2)}((O_s)_s).$$

For a partial matching N of size h between $\{0,\ldots,j+1\}$ and $\{0,\ldots,r+1\}$, let T_N be the merged coordinate set and let ι_L, ι_R send left and right coordinates into T_N, with $\iota_L(q) = \iota_R(s)$ exactly for matched pairs. Define

$$\mathcal{M}_{j,r}(h) := \{N : |N| = h\}, \qquad \eta_N := \frac{(m)_{(j+2)+(r+2)-h}}{(m)_{j+2}(m)_{r+2}},$$

and

$$\Gamma_N := \int \left\{ \prod_{q=0}^{j+1} L_q(O_{\iota_L(q)}) \right\} \left\{ \prod_{s=0}^{r+1} R_s(O_{\iota_R(s)}) \right\} dP_O^{\otimes T_N}((O_t)_{t\in T_N}).$$

The ordered-factorial covariance expansion gives

$$\mathrm{Cov}_{Q_P^{(m)}}(U_{k,a,j}, U_{\ell,b,r}) = (\eta_\varnothing - 1)\mu_{k,a,j}\mu_{\ell,b,r} + \sum_{h=1}^{\min(j+2,r+2)} \sum_{N\in\mathcal{M}_{j,r}(h)} \eta_N \Gamma_N.$$

Fix $N \in \mathcal{M}_{j,r}(h)$ with $h > 0$. Choose a matched pair (q,s). At the merged coordinate $\iota_L(q) = \iota_R(s)$, the factor L_q contains $\mathbf{1}\{X = k\}$ and the factor R_s contains $\mathbf{1}\{X = \ell\}$. Since $k \neq \ell$, their product is zero pointwise:

$$\mathbf{1}\{X = k\}\mathbf{1}\{X = \ell\} = 0.$$

Thus the merged kernel is identically zero and

$$\Gamma_N = 0 \qquad (h > 0).$$

All positive-size matching terms in the expansion therefore vanish.

It remains to bound the empty-matching term. The admissible-order and block-size assumptions give

$$|\eta_\varnothing - 1| \le \frac{2(K+2)^2}{m}.$$

For the product means, each one-coordinate mean is bounded by the corresponding cell mass:

$$\left| \int L_q(O)\, dP_O(O) \right| \le p_k, \qquad \left| \int R_s(O)\, dP_O(O) \right| \le p_\ell.$$

For the outcome-marked coordinate, if the relevant cell mass is zero then the arm-cell event is null and the marked integral is zero. If the relevant cell x has positive mass, Assumption 4 and $M \ge 1$ from Definition 1 give, for the corresponding arm c,

$$\left| \int y/M\, dP_{Y|A=c,X=x}(y) \right| \le \frac{1}{2},$$

so multiplying by the arm-cell probability bounds the marked coordinate by one half of the cell mass. The selector-only coordinates integrate to $P_O(X = x)$ or $P_O(X = x, A = c)$, both bounded by the cell mass. Since the product-law means factor coordinatewise,

$$|\mu_{k,a,j}| \le p_k^{j+2}, \qquad |\mu_{\ell,b,r}| \le p_\ell^{r+2}.$$

Combining the vanished positive-size terms with the preceding bounds yields

$$\left|\mathrm{Cov}_{Q_P^{(m)}}(U_{k,a,j}, U_{\ell,b,r})\right| \le \frac{2(K+2)^2}{m}\, p_k^{j+2} p_\ell^{r+2},$$

which is the claimed bound. □

Together, Lemmas 8 and 9 supply the weak-moment covariance control needed for the continuous-outcome polynomial lift. The same-cell inequality carries the additional term that remains after the two factorial arrays overlap in a cell; across distinct cells, the block-size term is the relevant covariance contribution. These estimates use the ordered distinct-index structure familiar from symmetric-statistic calculations [Hoeffding, 1948, Serfling, 1980].

For the heavy-cell part of a fixed polynomial branch, the argument uses empirical arm-cell means with a total value when an arm-cell count is empty. The convention is local to the block calculation and agrees with the totalized plug-in convention used in Algorithm 1.

Definition 11. [`synth_6`] (Empirical arm-cell mean $\widehat{\mu}_{ak}$) *For a block of observations $O_i = (X_i, A_i, Y_i)$, define the arm-cell count and outcome total by*

$$N_{ak} := \sum_i \mathbf{1}\{X_i = k, A_i = a\}, \qquad S_{ak} := \sum_i \mathbf{1}\{X_i = k, A_i = a\} Y_i.$$

The empirical arm-cell mean is

$$\widehat{\mu}_{ak} := \begin{cases} S_{ak}/N_{ak}, & N_{ak} > 0, \\ 0, & N_{ak} = 0. \end{cases}$$

The next aggregate covariance bound combines the coefficient envelope with the same-cell and distinct-cell moment bounds. It is the main stochastic input for summing the light-cell polynomial contributions over a deterministic set.

Lemma 10. [`lem:linear-mark-factorial-covariance`] (Factorial covariance bound) *For every overlap parameter $0 < \epsilon < 1/2$, there is a constant $C_\epsilon > 0$ such that the following statement holds. Let d, m, and K be nonnegative integers, let $M, \sigma, B \in \mathbb{R}$, let $P \in \mathcal{P}_{d,\epsilon,M,\sigma}$ as in Definition 1, and let $S \subseteq \{1, \dots, d\}$ be deterministic. For a block of m independent observations with joint law $Q_P^{(m)}$, write $V_i = Y_i/M$. For the polynomial degree K, define the shifted-Chebyshev coefficients*

$$g_{Kj} := (-1)^j \frac{2^{2j+3}}{K(K+j+2)} \binom{K+j+2}{2j+4}, \qquad 0 \le j \le K-2.$$

For $0 \le j \le K-2$, define

$$U_{k,a,j} := \frac{1}{(m)_{j+2}} \sum_{i_0,\dots,i_{j+1} \text{ distinct}} V_{i_0} \mathbf{1}\{X_{i_0} = k, A_{i_0} = a\} \mathbf{1}\{X_{i_1} = k\} \prod_{t=2}^{j+1} \mathbf{1}\{X_{i_t} = k, A_{i_t} = a\}.$$

Set

$$\widehat{P}_k := \sum_{j=0}^{K-2} \frac{g_{Kj}}{B^{j+1}}\{U_{k,1,j} - U_{k,0,j}\}.$$

Assume that

- ***(Degree.)*** $K \geq 2$.
- ***(Scale.)*** $B > 0$.
- ***(Light cells.)*** $p_k \leq B/4$ *for every* $k \in S$.
- ***(Block size.)*** $4(K+2)^2 \leq m$.
- ***(Shift calibration.)*** $4(K+2)/m \leq 3B/4$.

Then

$$\mathrm{Var}_{Q_P^{(m)}}\left(\sum_{k\in S} \widehat{P}_k\right) \leq \frac{C_\epsilon}{m} + C_\epsilon 6^{2K}\left\{dB^2 + \frac{d^2K^2B^2}{m}\right\}.$$

Proof of Lemma 10. Take $C_\epsilon = 64$. This constant is positive and depends only on ϵ.

Let

$$F_k := \widehat{P}_k = \sum_{j=0}^{K-2} \frac{g_{Kj}}{B^{j+1}}\{U_{k,1,j} - U_{k,0,j}\}.$$

All F_k have finite second moments under $Q_P^{(m)}$, since each is a finite linear combination of ordered one-mark factorial statistics and the conditional second-central-moment envelope in Definition 1 supplies the needed second moments.

Expanding the variance of the finite sum gives

$$\mathrm{Var}_{Q_P^{(m)}}\left(\sum_{k\in S} F_k\right) = \sum_{k\in S}\sum_{\ell\in S} \mathrm{Cov}_{Q_P^{(m)}}(F_k, F_\ell).$$

Consequently,

$$\mathrm{Var}_{Q_P^{(m)}}\left(\sum_{k\in S} F_k\right) \leq \sum_{k\in S}\sum_{\ell\in S} \left|\mathrm{Cov}_{Q_P^{(m)}}(F_k, F_\ell)\right|.$$

Fix $k \in S$, and write $p_k = P(X = k)$. The light-cell condition and the shift condition give, for every $0 \leq r \leq K-2$,

$$0 \leq p_k \leq B, \qquad \frac{5}{4}p_k \leq B, \qquad \frac{5}{4}p_k + \frac{r+2}{m} \leq B.$$

Indeed, $p_k \leq B/4$, while $r+2 \leq K$ and $4(K+2)/m \leq 3B/4$ control the final summand. Also $4(K+2)^2 \leq m$ gives

$$\frac{2(K+2)^2}{m} \leq \frac{1}{2}.$$

For each arm pair and each $j, r \leq K-2$, Lemma 8 therefore yields

$$\begin{aligned}\left|\mathrm{Cov}_{Q_P^{(m)}}(U_{k,a,j}, U_{k,b,r})\right| \leq{}& \frac{1}{2}p_k^{j+2}p_k^{r+2} \\ &+ \exp(1)\left(\frac{5}{4}p_k\right)^{r+2}\left(\frac{5}{4}p_k + \frac{r+2}{m}\right)^{j+2}.\end{aligned}$$

Expanding F_k, taking absolute values, and summing over both arm indices decomposes the same-cell contribution into the empty-matching and positive-matching weighted sums. By Lemma 7, the empty-matching part is bounded by

$$4 \cdot \frac{1}{2} \left\{ \sum_{j=0}^{K-2} \left| \frac{g_{Kj}}{B^{j+1}} \right| p_k^{j+2} \right\} \left\{ \sum_{r=0}^{K-2} \left| \frac{g_{Kr}}{B^{r+1}} \right| p_k^{r+2} \right\} \leq 4 \cdot \frac{1}{2} \{B6^K\}^2.$$

For the positive-matching part, the same coefficient envelope applies once with $(5/4)p_k$ and, for each fixed r, once with $(5/4)p_k + (r+2)/m$; using $\exp(1) \leq 3$,

$$\begin{aligned} &4 \sum_{r=0}^{K-2} \left| \frac{g_{Kr}}{B^{r+1}} \right| \exp(1) \left(\frac{5}{4} p_k \right)^{r+2} \left\{ \sum_{j=0}^{K-2} \left| \frac{g_{Kj}}{B^{j+1}} \right| \left(\frac{5}{4} p_k + \frac{r+2}{m} \right)^{j+2} \right\} \\ &\qquad \leq 4 \cdot 3 \{B6^K\}^2. \end{aligned}$$

Thus

$$\left| \mathrm{Cov}_{Q_P^{(m)}}(F_k, F_k) \right| \leq \left(4 \cdot \frac{1}{2} + 4 \cdot 3 \right) \{B6^K\}^2 \leq 16 \{B6^K\}^2.$$

For distinct $k, \ell \in S$, Lemmas 7 and 9 give

$$\left| \mathrm{Cov}_{Q_P^{(m)}}(F_k, F_\ell) \right| \leq 32 \, \frac{K^2}{m} \{B6^K\}^2.$$

Here all positive partial matchings between different cells are incompatible, so the only contribution is the empty-matching normalization correction

$$\frac{2(K+2)^2}{m} p_k^{j+2} p_\ell^{r+2};$$

summing over both arms contributes the factor 4, and $K \geq 2$ gives $(K+2)^2 \leq 4K^2$.

Since $|S| \leq d$, summing the diagonal and off-diagonal bounds gives

$$\mathrm{Var}_{Q_P^{(m)}} \left(\sum_{k \in S} F_k \right) \leq 16 d \{B6^K\}^2 + 32 d^2 \frac{K^2}{m} \{B6^K\}^2.$$

Using

$$\{B6^K\}^2 = B^2 \, 6^{2K},$$

enlarging the two coefficients to 64, and adding the nonnegative term $64/m$, we obtain

$$\mathrm{Var}_{Q_P^{(m)}} \left(\sum_{k \in S} \widehat{P}_k \right) \leq \frac{64}{m} + 64 \, 6^{2K} \left\{ dB^2 + \frac{d^2 K^2 B^2}{m} \right\}.$$

This is the asserted bound with $C_\epsilon = 64$. □

Lemma 10 isolates the continuous-outcome step in the polynomial estimator. The conditional second central moment envelope in Definition 1 is enough to control the one-mark factorial array after normalization by M, and the final display separates the block sampling term from the two approximation-scale terms that depend on d, B, and K.

The fixed-branch risk bound adds the heavy-cell plug-in component to the light-cell polynomial component. The statement is deterministic in the heavy set H; in the estimator of Algorithm 1, the sample split allows this result to be applied conditionally on the pilot block.

Lemma 11. [`lem:polynomial-fixed-branch-risk`] (Fixed branch risk bound) *For every overlap parameter $0 < \epsilon < 1/2$, there is a constant $C_\epsilon > 0$ such that the following holds. Let $d, m, K \in \mathbb{N}$, let $M, \sigma, B, \beta \in \mathbb{R}$, let $P \in \mathcal{P}_{d,\epsilon,M,\sigma}$ as in Definition 1, and let $H \subseteq \{1, \dots, d\}$ be deterministic. For one block of observations $O_i = (X_i, A_i, Y_i)$, $i = 1, \dots, m$, with joint law $Q_P^{(m)}$, define*

$$\widehat{R}_H := \sum_{k \in H} \widehat{p}_k \{\widehat{\mu}_{1k} - \widehat{\mu}_{0k}\}, \qquad R_H(P) := \sum_{k \in H} p_k \tau_k,$$

where $\widehat{p}_k$ is the empirical mass of cell k on the block and $\widehat{\mu}_{ak}$ is the arm-cell sample mean on positive arm-cell counts, with the corresponding empty-cell contribution set to zero. Let $\widehat{P}_k$ be the light factorial-polynomial contribution at scale B and degree K from Algorithm 1, and define

$$\Delta_H := \frac{\widehat{R}_H - R_H(P)}{M} + \left\{ \sum_{k \in H^c} \widehat{P}_k - \frac{1}{M} \sum_{k \in H^c} p_k \tau_k \right\}.$$

Let $\mathcal{E}_{m,\epsilon,\beta}(H)$ denote the boundary-safe lower-mass missing envelope,

$$\mathcal{E}_{m,\epsilon,\beta}(H) = \begin{cases} \dfrac{|H|}{\{(m-2)\epsilon/2\}^2 \beta}, & \{(m-2)\epsilon/2\}^2 \beta > 0, \\ \sum_{k \in H} p_k, & \{(m-2)\epsilon/2\}^2 \beta \le 0. \end{cases}$$

Assume:

- ***(Degree.)*** *$2 \le K$.*
- ***(Scales.)*** *$B > 0$ and $\beta > 0$.*
- ***(Heavy band.)*** *$\beta \le p_k$ for every $k \in H$.*
- ***(Light scale.)*** *$p_k \le B/4$ for every $k \in H^c$.*
- ***(Block size.)*** *$4(K+2)^2 \le m$.*
- ***(Shift calibration.)*** *$4(K+2)/m \le 3B/4$.*

Then

$$\mathbb{E}_{Q_P^{(m)}}\left[\Delta_H^2\right] \le 8 \left\{ \frac{8 \sum_{k \in H} p_k}{(m \vee 1)\epsilon} + \frac{6}{m \vee 1} + 4\, \mathcal{E}_{m,\epsilon,\beta}(H)^2 \right\} + 2 \left\{ \frac{C_\epsilon}{m} + C_\epsilon 6^{2K} \left(dB^2 + \frac{d^2 K^2 B^2}{m} \right) + \left(\frac{|H^c| B}{\epsilon K^2} \right)^2 \right\}.$$

Proof of Lemma 11. Fix $\epsilon \in (0, 1/2)$, and take $C_\epsilon > 0$ from the covariance bound in Lemma 10; the same constant is used in the fixed light-set calculation below.

For a sample block z, decompose

$$\Delta_H(z) = H_m(z) + L_m(z),$$

where

$$H_m(z) := \frac{\widehat{R}_H(z) - R_H(P)}{M}$$

and

$$L_m(z) := \sum_{k \in H^c} \widehat{P}_k(z) - \frac{1}{M} \sum_{k \in H^c} p_k \tau_k.$$

This is the deterministic fixed-branch normalized error in the statement.

The heavy part satisfies

$$\mathbb{E}_{Q_P^{(m)}}[H_m^2] \le 4\left\{\frac{8\sum_{k\in H} p_k}{(m\vee 1)\epsilon} + \frac{6}{m\vee 1} + 4\,\mathcal{E}_{m,\epsilon,\beta}(H)^2\right\}.$$

The fixed marked-ratio bound is applied to the deterministic heavy set H; the category masses in that bound are the cell masses p_k, and the normalization by M cancels the factor M^2 using $M \ge 1$.

For the light part, let

$$\widehat{L}_S(z) := \sum_{k\in S} \widehat{P}_k(z), \qquad T_S(P) := \frac{1}{M}\sum_{k\in S} p_k\tau_k, \qquad S = H^{\mathrm{c}}.$$

The bias-variance identity for the square-integrable statistic $\widehat{L}_S$ gives

$$\mathbb{E}_{Q_P^{(m)}}[L_m^2] = \mathrm{Var}_{Q_P^{(m)}}(\widehat{L}_S) + \left\{\mathbb{E}_{Q_P^{(m)}}[\widehat{L}_S] - T_S(P)\right\}^2.$$

The variance term is bounded by Lemma 10:

$$\mathrm{Var}_{Q_P^{(m)}}(\widehat{L}_S) \le \frac{C_\epsilon}{m} + C_\epsilon 6^{2K}\left(dB^2 + \frac{d^2K^2B^2}{m}\right),$$

because $S = H^{\mathrm{c}}$ satisfies the light-cell condition $p_k \le B/4$, together with the stated degree, block-size, and shift-calibration conditions.

For the bias term, the block-size condition implies $K \le m$, so the exact marked-factorial expectation identity applies term by term. With $s_{1k} = p_k\pi_k$, $s_{0k} = p_k(1-\pi_k)$, and the coefficients g_{Kj} from Lemma 10, it gives

$$\mathbb{E}_{Q_P^{(m)}}[\widehat{L}_S] = \sum_{k\in S}\Bigg[p_k\frac{s_{1k}(\mu_{1k}/M)}{B}\sum_{j=0}^{K-2} g_{Kj}\left(\frac{s_{1k}}{B}\right)^j - p_k\frac{s_{0k}(\mu_{0k}/M)}{B}\sum_{j=0}^{K-2} g_{Kj}\left(\frac{s_{0k}}{B}\right)^j\Bigg].$$

Therefore $\mathbb{E}_{Q_P^{(m)}}[\widehat{L}_S] - T_S(P)$ is the negative of the sum over $k \in S$ of

$$\frac{p_k\tau_k}{M} - \left[p_k\frac{s_{1k}(\mu_{1k}/M)}{B}\sum_{j=0}^{K-2} g_{Kj}\left(\frac{s_{1k}}{B}\right)^j - p_k\frac{s_{0k}(\mu_{0k}/M)}{B}\sum_{j=0}^{K-2} g_{Kj}\left(\frac{s_{0k}}{B}\right)^j\right].$$

For $x \in [0,1]$, the shifted-Chebyshev expansion of the first-kind Chebyshev polynomial gives

$$T_K(1-2x) = 1 - 2K^2x + 2K^2x^2\sum_{j=0}^{K-2} g_{Kj}x^j.$$

Since $1-2x \in [-1,1]$, the standard Chebyshev bound gives $|T_K(1-2x)| \le 1$, hence $|1-T_K(1-2x)| \le 2$. Combining this with the preceding identity and using $K > 0$ yields the reciprocal approximation certificate

$$x\left|1 - x\sum_{j=0}^{K-2} g_{Kj}x^j\right| \le \frac{1}{K^2}.$$

For a supported cell, overlap gives $\epsilon p_k \le s_{ak}$, the propensity range gives $s_{ak} \le p_k$, the model normalization gives $|\mu_{ak}/M| \le 1/2$, and the light condition gives $p_k \le B$. Applying the preceding certificate at $x = s_{ak}/B$ yields, for each arm $a \in \{0,1\}$,

$$\left| p_k \frac{\mu_{ak}}{M} - p_k \frac{s_{ak}(\mu_{ak}/M)}{B} \sum_{j=0}^{K-2} g_{Kj} \left(\frac{s_{ak}}{B}\right)^j \right| \le \frac{B}{2\epsilon K^2}.$$

Combining the two arm bounds gives the cellwise treated-minus-control bias bound

$$\left| \frac{p_k \tau_k}{M} - \left[p_k \frac{s_{1k}(\mu_{1k}/M)}{B} \sum_{j=0}^{K-2} g_{Kj} \left(\frac{s_{1k}}{B}\right)^j - p_k \frac{s_{0k}(\mu_{0k}/M)}{B} \sum_{j=0}^{K-2} g_{Kj} \left(\frac{s_{0k}}{B}\right)^j \right] \right| \le \frac{B}{\epsilon K^2}.$$

When $p_k = 0$, both the target and the displayed population polynomial vanish, so the same bound holds. Summing these cellwise bounds over $S = H^{\mathrm{c}}$ gives

$$\left| \mathbb{E}_{Q_P^{(m)}}[\widehat{L}_S] - T_S(P) \right| \le \frac{|H^{\mathrm{c}}|B}{\epsilon K^2}.$$

Therefore

$$\mathbb{E}_{Q_P^{(m)}}[L_m^2] \le \frac{C_\epsilon}{m} + C_\epsilon 6^{2K} \left(dB^2 + \frac{d^2 K^2 B^2}{m} \right) + \left(\frac{|H^{\mathrm{c}}|B}{\epsilon K^2} \right)^2.$$

Finally, $(u+v)^2 \le 2u^2 + 2v^2$ gives

$$\mathbb{E}_{Q_P^{(m)}}[\Delta_H^2] \le 2\mathbb{E}_{Q_P^{(m)}}[H_m^2] + 2\mathbb{E}_{Q_P^{(m)}}[L_m^2],$$

and substituting the two bounds above is exactly the displayed inequality. □

Lemma 11 decomposes the normalized branch error into a heavy-cell plug-in part, a light-cell polynomial variance part, and an approximation remainder. The boundary-safe envelope records the cost of missing an arm within a heavy cell under the lower-mass band, while the final term $(|H^{\mathrm{c}}|B/(\epsilon K^2))^2$ is the polynomial approximation contribution.

We next record the full-sample occupancy bound for the collision estimator. The event defining crossed cells in Algorithm 2 ensures that the treated and control sample means entering the statistic are evaluated only when both arm-cell counts are positive.

Lemma 12. [`lem:continuous-occupancy-collision-upper-all-d`] (Continuous occupancy collision bound) *For every overlap parameter $0 < \epsilon < 1/2$, there is a constant $C_\epsilon > 0$ such that the following holds.*

- ***(Sample size and alphabet.)*** *The integers n and d satisfy $n \ge 1$ and $d \ge 1$.*
- ***(Model.)*** *The distribution P belongs to $\mathcal{P}_{d,\epsilon,M,\sigma}$ from Definition 1, with its corresponding bounds on M and σ.*
- ***(Estimator.)*** *The estimator is the clipped occupancy-weighted treated-control estimator T_n^{col} defined in Algorithm 2 with clipping level M.*

Then

$$E_{Q_P^{(n)}}\left[\left(T_n^{\mathrm{col}} - \tau(P) \right)^2 \right] \le C_\epsilon M^2 \left(\frac{1}{n} + \sigma^2 + \frac{d}{n^2} \right).$$

Proof of Lemma 12. Fix $\epsilon \in (0, 1/2)$. Let $b_\epsilon, B_\epsilon > 0$ be the constants in Lemma 6, and set

$$K_\epsilon := \frac{16}{\epsilon^2(1-\epsilon)}, \qquad C_\epsilon := 2K_\epsilon B_\epsilon + \frac{2K_\epsilon B_\epsilon + 4}{b_\epsilon} + 2.$$

These constants are positive. Now fix $n, d \geq 1$, real parameters M, σ, and $P \in \mathcal{P}_{d,\epsilon,M,\sigma}$ as in Definition 1; let $Q_P^{(n)}$ be the n-fold observed-data law.

Write the unclipped statistic in Algorithm 2 with its zero fallback as

$$\widetilde{T}_n := \begin{cases} \dfrac{1}{R_n} \displaystyle\sum_{k: I_k = 1} N_k(\overline{Y}_{1k} - \overline{Y}_{0k}), & R_n > 0, \\ 0, & R_n = 0, \end{cases}$$

and write its design-centered version, totalized in the same way, as

$$C_n := \begin{cases} \dfrac{1}{R_n} \displaystyle\sum_{k: I_k = 1} N_k(\mu_{1k} - \mu_{0k}), & R_n > 0, \\ 0, & R_n = 0. \end{cases}$$

On $\{R_n > 0\}$, the difference is the finite-design weighted sum of supported arm-cell residuals

$$\widetilde{T}_n - C_n = \sum_{k: I_k = 1} \sum_{i=1}^{n} \frac{N_k}{R_n} \left\{ \frac{\mathbf{1}\{X_i = k, A_i = 1\}}{N_{1k}} (Y_i - \mu_{1k}) - \frac{\mathbf{1}\{X_i = k, A_i = 0\}}{N_{0k}} (Y_i - \mu_{0k}) \right\},$$

with value 0 on $\{R_n = 0\}$. The coefficients in this display depend only on the finite design $((X_i, A_i))_{i=1}^n$. Expanding the square therefore gives products of two supported residuals multiplied by a product of finite-design coefficients. When the two sample coordinates are distinct, the product-law integral is zero after conditioning on the finite design of the other coordinates and integrating one coordinate over its arm-cell fibers; the centering used is exactly

$$\int_{\{X=k, A=a\}} (Y - \mu_{ak})\, dP = 0.$$

When the two factors use the same coordinate but different arm-cell labels, their supports are disjoint, so the product is pointwise zero. The remaining same-coordinate, same-label terms are bounded by the arm-cell second central moments from Assumption 6, and summing the coordinate indicators gives the arm-cell counts. Hence

$$\mathbb{E}_{Q_P^{(n)}} \left[(\widetilde{T}_n - C_n)^2 \right] \leq M^2\, \mathbb{E}_{Q_P^{(n)}} \left[\mathbf{1}\{R_n > 0\} \frac{1}{R_n^2} \sum_{k: I_k = 1} N_k^2 \left(\frac{1}{N_{1k}} + \frac{1}{N_{0k}} \right) \right].$$

It remains to compare the design factor in the last display with reciprocal usable occupancy. Push the sample to the finite design $((X_i, A_i))_{i=1}^n$. In each positive-mass cell, the overlap condition in Assumption 3 gives both arm probabilities at least ϵp_k, equivalently $\pi_k \in [\epsilon, 1-\epsilon]$. A cell-label configuration that places an observation in a zero-mass cell has zero joint design weight, so it contributes zero before any conditional arm probability is formed.

Fix a positive-mass cell k, a cell-label configuration, and the arm assignments outside k. Let $m = N_k$ for this fixed cell-label configuration, and let r be the usable count contributed by all other cells under the fixed outside arm assignments. For $m \le 1$, the cell cannot be crossed and its local variance and local reciprocal-occupancy share are both zero. For $m \ge 2$, summing only over the m arm assignments inside cell k gives the local finite-binomial comparison

$$\sum_{j=0}^{m} \binom{m}{j} \pi_k^j (1-\pi_k)^{m-j} \frac{m^2}{(r+m)^2} \mathbf{1}\{0 < j < m\} \left(\frac{1}{j} + \frac{1}{m-j} \right)$$
$$\le \frac{4}{\epsilon^2(1-\epsilon)} \sum_{j=0}^{m} \binom{m}{j} \pi_k^j (1-\pi_k)^{m-j} \frac{m}{(r+m)^2} \mathbf{1}\{0 < j < m\}.$$

The denominator $(r+m)^2$ is the squared global usable count that applies when cell k is crossed; it is kept inside the local term throughout the comparison. The displayed inequality follows from the binomial inverse-arm bound

$$\sum_{j=0}^{m} \binom{m}{j} \pi_k^j (1-\pi_k)^{m-j} \mathbf{1}\{0 < j < m\} \left(\frac{1}{j} + \frac{1}{m-j} \right) \le \frac{4}{(m+1)\epsilon}$$

and the interior-mass lower bound

$$2\epsilon(1-\epsilon) \le \sum_{j=0}^{m} \binom{m}{j} \pi_k^j (1-\pi_k)^{m-j} \mathbf{1}\{0 < j < m\}.$$

Summing the local comparison over cells and outside assignments gives the finite-design inequality with constant $4/\{\epsilon^2(1-\epsilon)\}$. We retain the weaker constant $K_\epsilon = 16/\{\epsilon^2(1-\epsilon)\}$, obtaining

$$\mathbb{E}_{Q_P^{(n)}} \left[\mathbf{1}\{R_n > 0\} \frac{1}{R_n^2} \sum_{k:I_k=1} N_k^2 \left(\frac{1}{N_{1k}} + \frac{1}{N_{0k}} \right) \right] \le K_\epsilon \, \mathbb{E}_{Q_P^{(n)}} \left[\mathbf{1}\{R_n > 0\} \frac{1}{R_n} \right].$$

Combining the last two displays yields

$$\mathbb{E}_{Q_P^{(n)}} \left[(\widetilde{T}_n - C_n)^2 \right] \le K_\epsilon M^2 \, \mathbb{E}_{Q_P^{(n)}} \left[\mathbf{1}\{R_n > 0\} \frac{1}{R_n} \right].$$

The centered design bias obeys

$$\mathbb{E}_{Q_P^{(n)}} \left[(C_n - \tau(P))^2 \right] \le M^2 \left\{ \sigma^2 + Q_P^{(n)}(R_n = 0) \right\}.$$

Indeed, because $Q_P^{(n)}$ is the product observed-data law, every empirically usable cell has positive population mass almost surely: if $p_k = 0$, then no observation falls in cell k almost surely. On this full-probability event and on $\{R_n > 0\}$,

$$C_n - \tau(P) = \frac{1}{R_n} \sum_{k:I_k=1} N_k \delta_k, \qquad \delta_k := \mu_{1k} - \mu_{0k} - \tau(P),$$

so $C_n - \tau(P)$ is a convex average of deviations for positive-mass cells. The homogeneity condition in Assumption 5 gives $|C_n - \tau(P)| \le \sigma M$ there. On $\{R_n = 0\}$, the totalized definition gives $C_n = 0$, and $|\tau(P)| \le M$ by Lemma 1.

Put
$$x := \frac{n \vee d}{n^2}.$$
Since $n \geq 1$, $x > 0$, and Lemma 6 gives
$$\mathbb{E}_{Q_P^{(n)}}\left[\mathbf{1}\{R_n > 0\}\frac{1}{R_n}\right] \leq B_\epsilon\{x + \exp(-b_\epsilon/x)\}, \qquad Q_P^{(n)}(R_n = 0) \leq 2\exp(-b_\epsilon/x).$$
The elementary inequality
$$\exp(-b_\epsilon/x) \leq b_\epsilon^{-1}x, \qquad x > 0,$$
follows from $y \leq e^y$ with $y = b_\epsilon/x$. Also
$$x = \frac{n \vee d}{n^2} \leq \frac{1}{n} + \frac{d}{n^2}.$$

The clipping interval contains $\tau(P)$ by Lemma 1, so clipping the inner statistic to $[-M, M]$ is a contraction of squared distance from $\tau(P)$. Together with $(u+v)^2 \leq 2u^2 + 2v^2$, the preceding bounds give
$$\mathbb{E}_{Q_P^{(n)}}\left[(T_n^{\mathrm{col}} - \tau(P))^2\right] \leq 2K_\epsilon B_\epsilon M^2\{x + \exp(-b_\epsilon/x)\} + 2M^2\{\sigma^2 + 2\exp(-b_\epsilon/x)\}.$$
Using the two preceding displays and the definition of C_ϵ,
$$\mathbb{E}_{Q_P^{(n)}}\left[(T_n^{\mathrm{col}} - \tau(P))^2\right] \leq C_\epsilon M^2\left(\frac{1}{n} + \sigma^2 + \frac{d}{n^2}\right),$$
which is the asserted bound. □

Lemma 12 supplies the collision branch uniformly over all alphabet sizes. The three terms have the same interpretation as in the main text: ordinary sampling variation, the radius-controlled cost of replacing uncrossed cell effects by crossed-cell information, and the sparse-cell occupancy contribution.

The polynomial branch is calibrated by choosing the degree and the active range so that the fixed-branch inequality can be summed over the pilot-induced heavy-light split. The next result packages that calibration for the total estimator in Algorithm 1.

Lemma 13. [`lem:continuous-ratio-polynomial-upper-all-d`] (All-alphabet polynomial upper bound) *For every overlap parameter $0 < \epsilon < 1/2$, there exist a constant $C_\epsilon > 0$ and a polynomial handle, consisting of a cutoff $N_\epsilon \in \mathbb{N}$ and an active-range multiplier $\rho_\epsilon > 0$ such that*
$$N_\epsilon \leq n, \qquad d \leq \rho_\epsilon n \log(en) \quad \Longrightarrow \quad 2 \leq K_n$$
for all integers $n, d \geq 1$, with the following properties.

- ***(Complexity.)*** *The handle admits a constant $C_{\mathrm{op}} \in \mathbb{N}$, independent of n, d, and M, such that for every n, d and M the post-aggregation polynomial program for T_n^{poly} has at most*
$$C_{\mathrm{op}}(d+1)(K_n+1)^2$$
arithmetic operations.

- ***(Total bounded estimator.)*** *For every pair of integers $n, d \geq 1$ and every $M \geq 1$, the estimator T_n^{poly} determined by this handle is a measurable function of the n-sample observed data and satisfies*
$$T_n^{\mathrm{poly}}(s) \in [-M, M]$$
for every sample s.

- ***(Risk bound.)*** *For every pair of integers $n, d \geq 1$, every $M \geq 1$, every $0 \leq \sigma \leq 2$, and every $P \in \mathcal{P}_{d,\epsilon,M,\sigma}$ as in Definition 1,*
$$\mathbb{E}_{Q_P^{(n)}}\left[\{T_n^{\mathrm{poly}} - \tau(P)\}^2\right] \leq C_\epsilon M^2 \left\{\frac{1}{n} + \min\left(1, \frac{d^2}{n^2 \log^2(en)}\right)\right\}.$$

Proof of Lemma 13. Fix $0 < \epsilon < 1/2$. Lemma 11 supplies a constant $C_\epsilon^{\mathrm{fix}} > 0$ for every deterministic heavy set H satisfying the heavy-band and light-scale eligibility inequalities.

Choose N_ϵ large enough that, for all $n \geq N_\epsilon$,
$$8 \leq n, \qquad 2 \leq K_n, \qquad 240 \leq L := \log(en).$$

Set $\rho_\epsilon = 1$. The pair $(N_\epsilon, \rho_\epsilon)$ is a polynomial handle in the sense of Algorithm 1: whenever $0 < n$, $0 < d$, $N_\epsilon \leq n$, and $d \leq \rho_\epsilon n L$, the handle certifies $2 \leq K_n$.

For every $n, d \geq 1$ and $M \geq 1$, the estimator determined by this handle is measurable and takes values in $[-M, M]$.

The post-aggregation complexity certificate uses the aggregate table associated with Algorithm 1: pilot cell counts, estimation-block arm counts, and estimation-block arm outcome sums. Consider arithmetic programs over this table in which constants and table reads have zero structural cost; arithmetic and comparison nodes have unit cost; a cell sum contributes d accumulator operations plus the costs of its cell bodies; an r-term polynomial-order sum contributes r accumulator operations plus the costs of its term bodies; a falling-factorial primitive of order q has cost q; and a cell-subtraction primitive has cost one. For every n, d and M, there is such a post-aggregation program $A_{n,d,M}$ computing the handle-indexed estimator on every sample and satisfying
$$\mathrm{op}(A_{n,d,M}) \leq 128(d + 1)(K_n + 1)^2.$$
On the active branch, this program evaluates the heavy contribution, the Chebyshev light sums, the cell sum, and the clipping operation from the aggregate table; on the inactive branch, it is the constant-zero program. Thus the complexity item holds with $C_{\mathrm{op}} = 128$.

Now work on the active range $N_\epsilon \leq n$ and $d \leq nL$, and fix $P \in \mathcal{P}_{d,\epsilon,M,\sigma}$ as in Definition 1. With $m_1 = n - \lfloor n/2 \rfloor$ and $B = 4096L/m_1$, condition on the pilot block. The good pilot event is the simultaneous eligibility event
$$\begin{cases} 128L/|I_0| \leq p_k, & k \in \widehat{\mathcal{H}}, \\ p_k \leq 512L/|I_0|, & k \in \widehat{\mathcal{L}}. \end{cases}$$

On this event, the realized set $H = \widehat{\mathcal{H}}$ is eligible with lower band $128L/|I_0|$ and upper band $512L/|I_0|$. Since $n \geq 2$, the calibration gives $512L/|I_0| \leq B/4$, so the light-scale hypothesis of Lemma 11 holds for every $k \notin H$. The same active-range calibration gives
$$4(K_n + 2)^2 \leq m_1, \qquad \frac{4(K_n + 2)}{m_1} \leq \frac{3B}{4}.$$

Applying Lemma 11 conditionally on the pilot block, then using the uniform cardinality, total-mass, and missing-envelope bounds built into the fixed-branch simplification, gives for every eligible deterministic branch

$$\mathbb{E}\big[\Delta_H^2 \mid I_0\big] \le R^{\mathrm{br}}_{n,d},$$

where

$$R^{\mathrm{br}}_{n,d} = 8\left\{\frac{8}{m_1\epsilon} + \frac{6}{m_1} + 4\left(\frac{d}{\{((m_1-2)/2)\epsilon\}^2(128L/|I_0|)}\right)^2\right\} \\ + 2\left\{\frac{C^{\mathrm{fix}}_\epsilon}{m_1} + C^{\mathrm{fix}}_\epsilon 6^{2K_n}\left(dB^2 + \frac{d^2K_n^2B^2}{m_1}\right) + \left(\frac{dB}{\epsilon K_n^2}\right)^2\right\}.$$

The rate algebra uses the concrete hypotheses $8 \le n$, $2 \le K_n$, and $240 \le L$ to bound this normalized deterministic-branch expression by

$$R^{\mathrm{br}}_{n,d} \le D_\epsilon\left\{\frac{1}{n} + u_{n,d}\right\}, \qquad u_{n,d} := \frac{d^2}{n^2L^2},$$

with

$$D_\epsilon := \frac{128}{\epsilon} + 96 + \frac{2}{\epsilon^4} + 4C^{\mathrm{fix}}_\epsilon + 6C^{\mathrm{fix}}_\epsilon 8192^2 + 2\left(\frac{32768}{\epsilon\alpha_0^2}\right)^2.$$

At the calibrated pilot bands, the bad-selector probability satisfies

$$\mathbb{P}\{\text{the displayed pilot eligibility event fails}\} \le 2d\exp(-32L).$$

The finite-selector inequality for the clipped normalized statistic therefore gives a normalized stream-risk bound with penalty $4 \cdot 2d\exp(-32L) = 8d\exp(-32L)$. Since the active range has $d \le nL$ and the cutoff gives $240 \le L$,

$$8d\exp(-32L) \le \frac{1}{n}.$$

Thus the clipped normalized active-branch risk is bounded by

$$(D_\epsilon + 1)\left\{\frac{1}{n} + u_{n,d}\right\}.$$

Rescaling the clipped normalized bound by M^2 gives

$$\mathrm{MSE}_P(T_n^{\mathrm{poly}}) \le M^2(D_\epsilon + 1)\left\{\frac{1}{n} + u_{n,d}\right\}.$$

On the active range with $\rho_\epsilon = 1$, $d \le nL$ implies $u_{n,d} \le 1$, so this is the stated active-range bound with $\min\{1, u_{n,d}\} = u_{n,d}$.

It remains to cover the inactive branch of Algorithm 1. There the estimator is the zero fallback and Definition 1 gives $|\tau(P)| \le M$, hence its mean-squared error is at most M^2. The fallback rate certificate splits the inactive condition into the two cases used by the handle: if $n < N_\epsilon$, the finite cutoff makes a constant multiple of $1/n$ at least one; if $d > \rho_\epsilon nL$, then $\min\{1, u_{n,d}\}$ is bounded below by $\min\{1, \rho_\epsilon^2\}$. A constant $C^{\mathrm{fb}}_\epsilon > 0$ therefore satisfies

$$M^2 \le C^{\mathrm{fb}}_\epsilon M^2\left\{\frac{1}{n} + \min\{1, u_{n,d}\}\right\}$$

throughout the inactive branch. Taking the final risk constant to dominate both C^{fb}_ϵ and $D_\epsilon + 1$ proves the all-alphabet risk bound. □

Lemma 13 turns the light-cell analysis into a usable estimator bound. The zero fallback in the inactive range keeps the estimator total for every n, d, and the displayed risk bound gives the polynomial scale used by the selector in Algorithm 3. The operation count records that, after aggregation by cell and order, the implementation grows polynomially in the alphabet size and degree.

The finite-range construction theorem combines the polynomial and collision branches before the all-alphabet packaging in Theorem 1. It keeps the calibrated active range explicit, which is useful for downstream statements that invoke the original range condition.

Theorem 6. [`thm:robust-upper-construction-resolution`] (Robust upper construction) *For every overlap parameter $0 < \epsilon < 1/2$, there exist constants $C_\epsilon, c_\epsilon > 0$ and a calibrated polynomial handle consisting of a cutoff N and an active-range multiplier $\rho > 0$ such that, for every $n, d \geq 1$,*

$$N \leq n, \qquad d \leq \rho n \log(en)$$

implies that the associated polynomial degree is at least 2. For every $n, d \geq 1$, $M \geq 1$, and $0 \leq \sigma \leq 2$, the handle-indexed polynomial estimator T_n^{poly} and the occupancy-weighted estimator T_n^{col} of Algorithm 2 are measurable functions of the sample and take values in $[-M, M]$.

If, in addition,

$$d \leq c_\epsilon \frac{n^2}{\log(en)},$$

then for every $P \in \mathcal{P}_{d,\epsilon,M,\sigma}$ as defined in Definition 1,

$$E_{Q_P^{(n)}}\left[\left(T_n^{\mathrm{poly}} - \tau(P)\right)^2\right] \leq C_\epsilon M^2 \left\{\frac{1}{n} + \min\left(1, \frac{d^2}{n^2 \log^2(en)}\right)\right\},$$

and

$$E_{Q_P^{(n)}}\left[\left(T_n^{\mathrm{col}} - \tau(P)\right)^2\right] \leq C_\epsilon M^2 \left(\frac{1}{n} + \sigma^2 + \frac{d}{n^2}\right).$$

Theorem 6 states the two estimators on the same calibrated range. The polynomial branch contributes the large-alphabet approximation rate, and the occupancy branch contributes the radius-sensitive collision rate. Since both estimators are clipped into $[-M, M]$, the selector can compare their deterministic benchmark terms within the admissible class of Definition 5.

The final result in this appendix is the restricted selector bound. It is the finite-range version of the selector statement used in the main text and follows by applying the branch with the smaller displayed nonparametric benchmark.

Theorem 7. [`thm:frontier-upper`] (Restricted frontier upper bound) *For every overlap parameter ϵ with $0 < \epsilon < 1/2$, there exist constants $C_\epsilon, c_\epsilon > 0$ and a polynomial calibration handle (N, ρ) with $\rho > 0$ such that, for all positive integers n, d,*

$$N \leq n \quad \text{and} \quad d \leq \rho n \log(en) \quad \Longrightarrow \quad 2 \leq K_n,$$

where K_n is the active polynomial degree in the polynomial estimator. For every n, d, M, σ, assume:

- ***(Sample and alphabet.)*** *$n \geq 1$ and $d \geq 1$.*
- ***(Envelope and radius.)*** *$M \geq 1$ and $0 \leq \sigma \leq 2$.*

- ***(Restricted dimension.)***

$$d \le c_\epsilon \frac{n^2}{\log(en)}.$$

Let $T_n^{\mathrm{poly}} \in \mathcal{T}_{n,M}$ be the polynomial estimator determined by this handle, let $T_n^{\mathrm{col}} \in \mathcal{T}_{n,M}$ be the collision estimator, and let $\widehat{\tau}_n^\star$ be the known-radius selector of Algorithm 3 applied to T_n^{poly}, T_n^{col}, and σ. Put

$$r_{n,d,\sigma} = \frac{1}{n} + \min\left\{1,\ \min\left\{\frac{d^2}{n^2\log^2(en)},\ \sigma^2 + \frac{d}{n^2}\right\}\right\}.$$

Then $\widehat{\tau}_n^\star \in \mathcal{T}_{n,M}$, and for every $P \in \mathcal{P}_{d,\epsilon,M,\sigma}$ in Definition 1,

$$E_{Q_P^{(n)}}\left[(\widehat{\tau}_n^\star - \tau(P))^2\right] \le C_\epsilon M^2 r_{n,d,\sigma}.$$

Theorem 7 records the finite-range selector guarantee behind the all-alphabet statement in Theorem 2. The benchmark $r_{n,d,\sigma}$ combines the parametric term with the smallest of the clipped-zero, polynomial, and collision remainders. This is the constructive upper-bound input used when the main bracket is assembled in Theorem 4.

C Proofs for lower bounds and regime algebra

This appendix records the auxiliary lower-bound and algebraic statements that support the main bracket. The proof has four mechanisms. The parametric lower bound supplies the M^2/n sampling term. The exact-homogeneity transport supplies the capped collision baseline $M^2\{n^{-1} + \min(1, d/n^2)\}$. The radius-channel transport supplies the rare-cell term $M^2\sigma^2\min\{1, u_{n,d}\}$. The final rate algebra compares these lower benchmarks with the selector benchmark and identifies the matched regimes and residual wedge.

Lemma 14. [`lem:parametric-lower`] (Parametric minimax lower bound) *For every overlap parameter ϵ satisfying $0 < \epsilon < 1/2$, there exists a constant $c_\epsilon > 0$ such that the following holds. For every $n, d \in \mathbb{N}$ and every $M, \sigma \in \mathbb{R}$, suppose that*

- ***(Sample and alphabet sizes.)*** *$n \ge 1$ and $d \ge 1$;*
- ***(Outcome radius.)*** *$M \ge 1$;*
- ***(Noise radius.)*** *$0 \le \sigma \le 2$.*

Then the minimax mean-squared risk $\mathsf{R}_{n,d,\epsilon,M,\sigma}$ of Definition 5 satisfies

$$\mathsf{R}_{n,d,\epsilon,M,\sigma} \ge \frac{c_\epsilon M^2}{n}.$$

Proof of Lemma 14. Take

$$c_\epsilon := \frac{1}{100}.$$

Fix n, d, M, σ satisfying the displayed hypotheses, and choose a cell $k \in [d]$. Put

$$\delta := \frac{2/5}{\sqrt{n}}, \qquad u_v := M(1/2 + v\delta - 1/2) \quad (v \in \{0,1\}).$$

Thus $u_0 = 0$ and $u_1 = M\delta$. Moreover,

$$\delta^2 = \frac{4}{25n}, \qquad 0 \le \delta \le \frac{2}{5}, \qquad |u_0| \le M/2, \qquad |u_1| \le M/2.$$

For $v \in \{0,1\}$, let P_v be the one-cell test law with $X = k$, $P(A = 1 \mid X = k) = 1/2$, control potential outcome $Y(0) = 0$, and treated potential outcome $Y(1)$ distributed on $\{-M/2, M/2\}$ with mean u_v. Equivalently, draw A with probability $1/2$ independently of this treated endpoint and set $Y = Y(A)$. Thus, for every cell ℓ, arm a, and measurable $S_0, S_1 \subseteq \mathbb{R}$,

$$P_v\{X = \ell, A = a, Y(0) \in S_0, Y(1) \in S_1\}\, P_v\{X = \ell\} = P_v\{X = \ell, A = a\}\, P_v\{X = \ell, Y(0) \in S_0, Y(1) \in S_1\},$$

so the conditional exchangeability requirement in Definition 1 holds, and consistency holds from $Y = Y(A)$. The same construction has cell mass one at k, overlap $1/2$, conditional means 0 and u_v, second central moments bounded by $(M/2)^2 \le M^2$, and cell deviation 0. Together with $0 < \epsilon < 1/2$, $M \ge 1$, and $0 \le \sigma \le 2$, this gives

$$P_v \in \mathcal{P}_{d,\epsilon,M,\sigma} \qquad (v = 0, 1).$$

The targets are

$$\tau(P_0) = 0, \qquad \tau(P_1) = M\delta.$$

Introduce the binary one-cell endpoint pair used for the comparison, and write $\widetilde{Q}_v^{(m)}$ for its m-fold binary observed law under $v \in \{0,1\}$. Both source laws have the unique cell, propensity $m_0 = 1/2$, and control endpoint probability $g_0 = 1/2$. The null source law has treated endpoint probability $g_1 = 1/2$, while the perturbed source law has treated endpoint probability $g_1 + \delta = 1/2 + \delta$; this is admissible because $0 \le \delta$ and $1/2 + \delta \le 1$. The one-observation chi-square calculation gives

$$\chi^2\Big(\widetilde{Q}_1^{(1)}, \widetilde{Q}_0^{(1)}\Big) = \frac{(1/2)\delta^2}{(1/2)(1 - 1/2)}.$$

The null one-observation atoms are positive, and tensorization gives

$$1 + \chi^2\Big(\widetilde{Q}_1^{(n)}, \widetilde{Q}_0^{(n)}\Big) = \left(1 + \frac{(1/2)\delta^2}{(1/2)(1 - 1/2)}\right)^n \le \exp\left(n \frac{(1/2)\delta^2}{(1/2)(1 - 1/2)}\right).$$

Since

$$n\, \frac{(1/2)\delta^2}{(1/2)(1 - 1/2)} = \frac{8}{25} \le \log 2,$$

we have $\chi^2(\widetilde{Q}_1^{(n)}, \widetilde{Q}_0^{(n)}) \le 1$. The chi-square-to-total-variation inequality and symmetry of total variation therefore yield

$$\mathrm{TV}\Big(\widetilde{Q}_0^{(n)}, \widetilde{Q}_1^{(n)}\Big) \le \frac{1}{2}\sqrt{\chi^2\Big(\widetilde{Q}_1^{(n)}, \widetilde{Q}_0^{(n)}\Big)} \le \frac{1}{2}.$$

The two real observed n-sample laws are the corresponding images under the coordinatewise affine map

$$(x, a, b) \longmapsto (k, a,\ \mathbf{1}\{a = 1\} M(b - 1/2)).$$

Therefore, by data processing for total variation,

$$\mathrm{TV}\Big(Q_{P_0}^{(n)}, Q_{P_1}^{(n)}\Big) \le \frac{1}{2}.$$

For any measurable estimator $T \in \mathcal{T}_{n,M}$, the two-point testing inequality gives

$$\frac{(M\delta)^2}{16} \leq \max_{v\in\{0,1\}} E_{Q^{(n)}_{P_v}}\left[(T-\tau(P_v))^2\right].$$

The supremum over $\mathcal{P}_{d,\epsilon,M,\sigma}$ is finite because $T \in [-M, M]$ and $|\tau(P)| \leq M$ by Lemma 1, so the preceding bound passes through the estimator infimum in Definition 5. Using $\delta^2 = 4/(25n)$,

$$\mathsf{R}_{n,d,\epsilon,M,\sigma} \geq \frac{M^2\delta^2}{16} = \frac{M^2}{100n} = \frac{c_\epsilon M^2}{n}.$$

□

Lemma 14 anchors the lower benchmark at the scale of estimating a bounded real mean from n observations. This component is combined below with the exact-homogeneity collision scale and the radius-channel rare-cell scale.

The radius-sensitive converse is organized through a channel from binary source experiments into the real-outcome class. The channel preserves the cell and treatment-assignment structure, contracts binary response contrasts by $\sigma/2$, and then rescales by M.

Definition 12. [`synth_5`] (Radius-channel hard-family embedding) *For $0 \leq \sigma \leq 2$, set $\lambda := \sigma/2$. The embedding Ψ sends a binary source law from the radius-channel hard family, after padding unused cells with zero mass in the d-alphabet, to the real-outcome law obtained by the hypothesis-independent channel*

$$B(a) \longmapsto B'(a) \sim \text{Bernoulli}\left\{\frac{1}{2} + \lambda\left(B(a) - \frac{1}{2}\right)\right\}, \qquad Y(a) := M\{B'(a) - 1/2\}, \qquad Y = Y(A).$$

For source laws P_0 and P_1 in the realized family, the embedded targets satisfy

$$\tau(\Psi(P_1)) - \tau(\Psi(P_0)) = M\frac{\sigma}{2}\{\psi(P_1) - \psi(P_0)\}.$$

Definition 12 is the finite-sample transport map used in the radius-channel lower bound. The target identity shows that a binary separation for ψ becomes an average-treatment-effect separation with scale $M\sigma$ in the embedded real-outcome experiment.

The next two transport lemmas calibrate the capped alphabet sizes for the exact-homogeneity and radius-channel components. Their role is purely quantitative: they convert the published binary lower-risk levels into the normalized expressions used by the finite-range bracket.

Lemma 15. [`lem:capped-exact-transport-package`] (Capped exact transport scale) *For every overlap level $0 < \epsilon < 1/2$, there exist constants $b^{\text{ex}}_\epsilon, c^{\text{ex}}_\epsilon > 0$ and a cutoff $N^{\text{ex}}_\epsilon \in \mathbb{N}$ such that the following properties hold.*

- ***(Capped alphabet.)*** *For every pair of integers $n, d \geq 1$, define the exact capped alphabet size from Algorithm 4 by*

$$d_{\text{ex}} = \min\left\{d, \max\left\{1, \left\lfloor b^{\text{ex}}_\epsilon n^2\right\rfloor\right\}\right\}.$$

Then $0 < d_{\text{ex}} \leq d$, and whenever $N^{\text{ex}}_\epsilon \leq n$,

$$d_{\text{ex}} \leq b^{\text{ex}}_\epsilon n^2.$$

- ***(Estimator-wise source risk.)*** *For every $n, d \geq 1$, with d_{ex} defined above, there is a real number $L_{\mathrm{ex}} = L_{\mathrm{ex}}(n, d)$ such that every measurable estimator*

$$\widehat{\psi} : \big(\{1, \dots, n\} \to \big(\{1, \dots, d_{\mathrm{ex}}\} \times \{0,1\} \times \{0,1\}\big)\big) \to \mathbb{R}$$

admits an exactly homogeneous binary law R over d_{ex} cells at overlap ϵ for which

$$L_{\mathrm{ex}} \leq \mathbb{E}_{Q_R^{(n)}}\left[\{\widehat{\psi} - \psi(R)\}^2\right].$$

- ***(Transport normalization.)*** *For every $n, d \geq 1$, with the same $L_{\mathrm{ex}}(n, d)$, and for every $M \in \mathbb{R}$,*

$$c_\epsilon^{\mathrm{ex}} M^2 \left\{\frac{1}{n} + \min\left(1, \frac{d}{n^2}\right)\right\} \leq M^2 L_{\mathrm{ex}}.$$

Proof of Lemma 15. Fix $\epsilon \in (0, 1/2)$. Let $a_\epsilon, b_\epsilon > 0$ and $N_\epsilon \in \mathbb{N}$ be the constants in Lemma 4. Set

$$b_\epsilon^{\mathrm{ex}} := \min\{b_\epsilon, 1/2\}, \qquad K_\epsilon := 1 + N_\epsilon + \frac{1}{b_\epsilon^{\mathrm{ex}}},$$

$$c_\epsilon^{\mathrm{ex}} := \min\left\{\frac{a_\epsilon b_\epsilon^{\mathrm{ex}}}{4}, \frac{1}{200 K_\epsilon}\right\}, \qquad N_\epsilon^{\mathrm{ex}} := \max\left\{N_\epsilon, \left\lceil \frac{1}{b_\epsilon^{\mathrm{ex}}} \right\rceil\right\}.$$

The two real constants have the asserted signs, and N_ϵ^{ex} is a natural cutoff.

For $n, d \geq 1$, put

$$x := b_\epsilon^{\mathrm{ex}} n^2, \qquad d_{\mathrm{ex}} := \min\{d, \max\{1, \lfloor x \rfloor\}\}.$$

Since $d \geq 1$, immediately

$$0 < d_{\mathrm{ex}} \leq d.$$

If $N_\epsilon^{\mathrm{ex}} \leq n$, then $1/b_\epsilon^{\mathrm{ex}} \leq n$, so $b_\epsilon^{\mathrm{ex}} n \geq 1$. Together with $n \geq 1$, this gives $1 \leq x$, and therefore

$$d_{\mathrm{ex}} \leq \max\{1, \lfloor x \rfloor\} = \lfloor x \rfloor \leq x = b_\epsilon^{\mathrm{ex}} n^2.$$

It remains to construct L_{ex}. First suppose

$$N_\epsilon \leq n \qquad \text{and} \qquad 1 \leq x.$$

Then

$$d_{\mathrm{ex}} = \min\{d, \lfloor x \rfloor\} \quad \text{and} \quad d_{\mathrm{ex}} \leq b_\epsilon n^2,$$

because $b_\epsilon^{\mathrm{ex}} \leq b_\epsilon$. Define

$$s_{\mathrm{ex}} := \frac{1}{n} + \frac{d_{\mathrm{ex}}}{n^2}, \qquad L_{\mathrm{ex}} := \frac{a_\epsilon}{2} s_{\mathrm{ex}}.$$

The quantity s_{ex} is positive. Hence L_{ex} lies strictly below $a_\epsilon s_{\mathrm{ex}}$, and Lemma 4 gives

$$a_\epsilon s_{\mathrm{ex}} \leq \inf_{\widehat{\psi}} \sup_{R \in \mathcal{H}^{\mathrm{bin}}_{n, d_{\mathrm{ex}}}(\epsilon)} \mathbb{E}_{Q_R^{(n)}}\left[\{\widehat{\psi} - \psi(R)\}^2\right].$$

Since L_{ex} is strictly below the displayed minimax lower bound, the definition of the infimum and supremum implies that every measurable estimator has some exactly homogeneous binary source law R over d_{ex} cells such that

$$L_{\mathrm{ex}} \leq \mathbb{E}_{Q_R^{(n)}}\left[\{\widehat{\psi} - \psi(R)\}^2\right].$$

In this large-source case,

$$\frac{b_\epsilon^{\mathrm{ex}}}{2}\left\{\frac{1}{n}+\min\left(1,\frac{d}{n^2}\right)\right\}\le \frac{1}{n}+\frac{d_{\mathrm{ex}}}{n^2}.$$

If $d\le \lfloor x\rfloor$, then $d_{\mathrm{ex}}=d$ and $d/n^2\le 1$. If $d>\lfloor x\rfloor$, then $d_{\mathrm{ex}}=\lfloor x\rfloor$ and $\lfloor x\rfloor\ge x/2$, so $d_{\mathrm{ex}}/n^2\ge b_\epsilon^{\mathrm{ex}}/2$; together with $b_\epsilon^{\mathrm{ex}}/2\le 1$, this absorbs both terms on the left.

Since $c_\epsilon^{\mathrm{ex}}\le a_\epsilon b_\epsilon^{\mathrm{ex}}/4$, the preceding display gives, for every $M\in\mathbb{R}$,

$$c_\epsilon^{\mathrm{ex}}M^2\left\{\frac{1}{n}+\min\left(1,\frac{d}{n^2}\right)\right\}\le M^2L_{\mathrm{ex}}.$$

Now consider the complementary case: either $n<N_\epsilon$ or $x<1$. Define

$$L_{\mathrm{ex}}:=\frac{1}{200n}.$$

Since $d_{\mathrm{ex}}>0$, the exact binary source class over d_{ex} cells contains the following two uniform-mass endpoint laws. In both laws the propensity is $1/2$ in every cell and the control-arm binary mean is $1/2$ in every cell. Under R_0, the treated-arm binary mean is $1/2$; under R_1, it is $1/2+\delta$, where

$$\delta:=\frac{2}{5\sqrt{n}},\qquad \delta^2=\frac{4}{25n}.$$

The inequality $n\ge 1$ gives $0<\delta\le 2/5$, so both endpoint laws are valid binary laws with overlap ϵ and a common cell effect. Their source targets satisfy

$$\psi(R_0)=0,\qquad \psi(R_1)=\delta.$$

Moreover

$$n\,\frac{(1/2)\delta^2}{(1/2)(1-1/2)}=\frac{8}{25}\le\log 2,$$

so the Bernoulli-product total-variation bound gives

$$\mathrm{TV}\Big(Q_{R_0}^{(n)},Q_{R_1}^{(n)}\Big)\le\frac{1}{2}.$$

The two-point testing bound therefore gives the exact-source minimax inequality

$$\frac{1}{100n}=\frac{\delta^2}{16}\le\inf_{\widehat{\psi}}\sup_{R\in\mathcal{H}^{\mathrm{bin}}_{n,d_{\mathrm{ex}}}(\epsilon)}\mathbb{E}_{Q_R^{(n)}}\left[\{\widehat{\psi}-\psi(R)\}^2\right].$$

Because $1/(200n)<1/(100n)$, the same infimum-supremum argument yields: every measurable estimator has some exactly homogeneous binary source law R over d_{ex} cells satisfying

$$L_{\mathrm{ex}}\le\mathbb{E}_{Q_R^{(n)}}\left[\{\widehat{\psi}-\psi(R)\}^2\right].$$

In this fallback case,

$$\frac{1}{n}+\min\left(1,\frac{d}{n^2}\right)\le\frac{K_\epsilon}{n}.$$

Indeed, if $n < N_\epsilon$, then $N_\epsilon/n \ge 1$; if $x < 1$, then $1/(b_\epsilon^{\mathrm{ex}} n) \ge 1$. In either case the right side

$$\frac{K_\epsilon}{n} = \frac{1}{n} + \frac{N_\epsilon}{n} + \frac{1}{b_\epsilon^{\mathrm{ex}} n}$$

dominates $n^{-1} + 1$, and the displayed bound follows.

Because $c_\epsilon^{\mathrm{ex}} \le 1/(200K_\epsilon)$, this gives

$$c_\epsilon^{\mathrm{ex}} M^2 \left\{ \frac{1}{n} + \min\left(1, \frac{d}{n^2}\right) \right\} \le M^2 \frac{1}{200n} = M^2 L_{\mathrm{ex}}.$$

The two cases supply the same capped alphabet properties, the estimator-wise source risk, and the transport normalization for all $n, d \ge 1$. □

The exact transport package expresses the binary exactly homogeneous collision lower bound at the scale of the real-outcome target. The capped alphabet d_{ex} keeps the embedded source within the available cells, while the normalization clause yields the baseline $M^2\{n^{-1} + \min(1, d/n^2)\}$.

Lemma 16. [`lem:capped-radial-transport-scale`] (Capped radial transport scale) *Let n, d be integers and let $a, b, M, \sigma \in \mathbb{R}$. Suppose that*

- ***(Sample size.)*** *$n \ge 3$.*
- ***(Alphabet size.)*** *$d \ge 1$.*
- ***(Scale.)*** *$a \ge 0$.*
- ***(Radial cap.)*** *$0 < b \le 4$.*
- ***(Range.)*** *$1 \le (b/2)n\log(en)$.*

Define the capped radial alphabet size from Algorithm 4 by

$$d_{\mathrm{rad}} := \min\left\{d,\ \max\left\{1, \lfloor (b/2)n\log(en)\rfloor\right\}\right\}.$$

Then

$$\frac{ab^2}{128} M^2\sigma^2 \min\left\{1, \frac{d^2}{n^2\log^2(en)}\right\} \le \left(\frac{M\sigma}{2}\right)^2 \frac{a}{2}\left\{\frac{1}{n} + \frac{d_{\mathrm{rad}}^2}{n^2\log^2 n}\right\}.$$

Proof of Lemma 16. Let

$$L := \log(en), \qquad x := \frac{b}{2}nL, \qquad d_{\mathrm{rad}} = \min\left\{d, \max\{1, \lfloor x\rfloor\}\right\}.$$

Since $n \ge 3$, both $\log n$ and L are positive, $L \ge \log n$, and the denominator nL is positive. The range assumption gives $1 \le x$, hence $\lfloor x\rfloor > 0$ and $\max\{1, \lfloor x\rfloor\} = \lfloor x\rfloor$.

First,

$$\frac{b}{4}\min\left\{1, \frac{d}{nL}\right\} \le \frac{d_{\mathrm{rad}}}{nL}.$$

If $d \le \lfloor x\rfloor$, then $d_{\mathrm{rad}} = d$. The fraction $d/(nL)$ is nonnegative, so $\min\{1, d/(nL)\}$ is nonnegative and at most $d/(nL)$. Together with $b/4 \le 1$, this gives

$$\frac{b}{4}\min\left\{1, \frac{d}{nL}\right\} \le \min\left\{1, \frac{d}{nL}\right\} \le \frac{d}{nL} = \frac{d_{\mathrm{rad}}}{nL}.$$

If $d > \lfloor x \rfloor$, then $d_{\text{rad}} = \lfloor x \rfloor$. Since $x \geq 1$, $\lfloor x \rfloor \geq x/2$, and positivity of nL gives

$$\frac{d_{\text{rad}}}{nL} \geq \frac{x/2}{nL} = \frac{b}{4}.$$

Also $\min\{1, d/(nL)\} \leq 1$ and $b/4 \geq 0$, so the desired inequality follows in this case as well.

Both sides of the preceding inequality are nonnegative, hence squaring gives

$$\left(\frac{b}{4} \min\left\{1, \frac{d}{nL}\right\}\right)^2 \leq \left(\frac{d_{\text{rad}}}{nL}\right)^2.$$

It remains to rewrite the squared minimum. Since $d/(nL) \geq 0$,

$$\left(\min\left\{1, \frac{d}{nL}\right\}\right)^2 = \min\left\{1, \frac{d^2}{n^2L^2}\right\}.$$

Indeed, if $d/(nL) \leq 1$, then the left minimum is $d/(nL)$, and nonnegativity gives $(d/(nL))^2 \leq 1$. If $1 \leq d/(nL)$, then the left minimum is 1, and nonnegativity gives $1 \leq (d/(nL))^2$. Using also $(d/(nL))^2 = d^2/(n^2L^2)$, the squared inequality becomes

$$\frac{b^2}{16} \min\left\{1, \frac{d^2}{n^2L^2}\right\} \leq \frac{d_{\text{rad}}^2}{n^2L^2}.$$

Since $L \geq \log n > 0$,

$$\frac{d_{\text{rad}}^2}{n^2L^2} \leq \frac{d_{\text{rad}}^2}{n^2 \log^2 n}.$$

Adding the nonnegative term $1/n$ on the right therefore yields

$$\frac{b^2}{16} \min\left\{1, \frac{d^2}{n^2 \log^2(en)}\right\} \leq \frac{1}{n} + \frac{d_{\text{rad}}^2}{n^2 \log^2 n}.$$

Multiplying this inequality by the nonnegative factor

$$\left(\frac{M\sigma}{2}\right)^2 \frac{a}{2}$$

and simplifying constants yields

$$\frac{ab^2}{128} M^2\sigma^2 \min\left\{1, \frac{d^2}{n^2 \log^2(en)}\right\} \leq \left(\frac{M\sigma}{2}\right)^2 \frac{a}{2} \left\{\frac{1}{n} + \frac{d_{\text{rad}}^2}{n^2 \log^2 n}\right\}.$$

□

Lemma 16 translates the source lower-risk scale into the rare-cell expression $M^2\sigma^2 \min\{1, d^2/(n^2 \log^2(en))\}$. The capped d_{rad} matches the radius-channel construction in Algorithm 4.

Small sample sizes are absorbed by a separate finite-sample comparison. This keeps the radius-channel lower bound calibrated uniformly over the range in which the asymptotic radial cap is inactive.

Lemma 17. [`lem:finite-sample-radial-transport-scale`] (Finite-sample radial scale) *Let $n, d, N \in \mathbb{N}$ satisfy $1 \le n \le N$, and let $M, \sigma \in \mathbb{R}$. Define*

$$u_{n,d} = \frac{d^2}{n^2 \log^2(en)}.$$

Then

$$\frac{1}{800N} M^2\sigma^2 \min\{1, u_{n,d}\} \le \left(\frac{M\sigma}{2}\right)^2 \frac{1}{200n}.$$

Proof of Lemma 17. Put

$$u_{n,d} = \frac{d^2}{n^2 \log^2(en)}.$$

The minimum is bounded by its first entry, so

$$\min\{1, u_{n,d}\} \le 1.$$

The assumptions $1 \le n \le N$ imply $N > 0$. Hence

$$0 \le \frac{1}{800N} M^2\sigma^2,$$

since the remaining factors are squares. Multiplying the preceding minimum bound by this nonnegative factor gives

$$\frac{1}{800N} M^2\sigma^2 \min\{1, u_{n,d}\} \le \frac{1}{800N} M^2\sigma^2.$$

The same assumptions give

$$\frac{1}{N} \le \frac{1}{n}.$$

Multiplying this reciprocal bound by the nonnegative factors $1/800$, M^2, and σ^2 yields

$$\frac{1}{800N} M^2\sigma^2 \le \frac{1}{800n} M^2\sigma^2.$$

Finally,

$$\frac{1}{800n} M^2\sigma^2 = \left(\frac{M\sigma}{2}\right)^2 \frac{1}{200n},$$

which proves the claim. □

Together, Lemmas 16 and 17 supply the scale algebra behind the finite-range radius-channel converse. The converse itself combines the transported radial component with the parametric and exact-homogeneity components.

Theorem 8. [`thm:radius-channel-converse`] (Radius-channel converse) *For every overlap parameter ϵ satisfying $0 < \epsilon < 1/2$, there exist constants c_ϵ and $b_\epsilon^{\mathrm{rad}}$ such that*

$$0 < c_\epsilon \le 1, \qquad 0 < b_\epsilon^{\mathrm{rad}}.$$

For every $n, d \in \mathbb{N}$, $M \in \mathbb{R}$, and $\sigma \in \mathbb{R}$, suppose that

- ***(Sample and alphabet.)*** *$0 < n$ and $0 < d$.*
- ***(Scale and radius.)*** *$1 \le M$ and $0 \le \sigma \le 2$.*
- ***(Restricted range.)***

$$d \le c_\epsilon \frac{n^2}{\log(en)}.$$

Then the minimax risk from Definition 5 satisfies

$$\mathsf{R}_{n,d,\epsilon,M,\sigma} \ge c_\epsilon M^2 \left\{ \frac{1}{n} + \frac{d}{n^2} + \sigma^2 \min\left(1, \frac{d^2}{n^2 \log^2(en)}\right) \right\}.$$

Moreover, the radius-channel hard family can be realized with the same constants: there are an integer cap, a source class, a padding map into the d-alphabet, an embedding Ψ of the source laws into real-outcome laws, and a coupling for the binary full-observation experiment such that the realized radial-channel data have radius parameter $b_\epsilon^{\mathrm{rad}}$, and for all source laws P_0 and P_1,

$$\tau(\Psi(P_1)) - \tau(\Psi(P_0)) = M \frac{\sigma}{2} \{\psi(P_1) - \psi(P_0)\}.$$

For every total measurable $[-M, M]$-valued estimator $\widehat{\tau}$, some source law P in this realized family satisfies

$$c_\epsilon M^2 \sigma^2 \min\left(1, \frac{d^2}{n^2 \log^2(en)}\right) \le \mathbb{E}_{\Psi(P)}\left[\{\widehat{\tau} - \tau(\Psi(P))\}^2\right].$$

Theorem 8 establishes the finite-range lower benchmark used in the original calibrated bracket. Its first two terms give the parametric and collision baseline, and the final term is the radius-channel contribution. The displayed source-target identity records the mechanism of the transport: source contrasts in ψ become real-outcome contrasts in τ through the attenuation $\sigma/2$ and the outcome scale M, as in classical two-point and fuzzy-hypothesis lower-bound arguments [Le Cam, 1986, Tsybakov, 2009, Cai and Low, 2011, Donoho et al., 1990].

Combining the finite-range converse with the constructive selector upper bound gives the restricted-range bracket. This statement is the calibrated predecessor of the all-alphabet bracket in Theorem 4.

Theorem 9. [`thm:two-sided-minimax-bracket`] (Two-sided minimax bracket) *For every overlap parameter ϵ with $0 < \epsilon < 1/2$, there exist real constants c_ϵ and C_ϵ such that*

$$0 < c_\epsilon \le 1, \qquad 1 \le C_\epsilon, \qquad c_\epsilon \le C_\epsilon.$$

For every pair of positive integers n, d and every $M, \sigma \in \mathbb{R}$, assume:

- ***(Scale.)*** *$1 \le M$.*
- ***(Radius.)*** *$0 \le \sigma \le 2$.*
- ***(Range.)***

$$d \le \frac{c_\epsilon n^2}{\log(en)}.$$

Let $\mathsf{R}_{n,d,\epsilon,M,\sigma}$ be the minimax mean-squared risk in Definition 5. Then

$$c_\epsilon M^2 \left\{\frac{1}{n} + \frac{d}{n^2} + \sigma^2 \min\left(1, \frac{d^2}{n^2 \log^2(en)}\right)\right\} \le \mathsf{R}_{n,d,\epsilon,M,\sigma} \le C_\epsilon M^2 \left\{\frac{1}{n} + \min\left(1, \min\left(\frac{d^2}{n^2 \log^2(en)}, \sigma^2 + \frac{d}{n^2}\right)\right)\right\}.$$

Theorem 9 places the finite-range upper and lower benchmarks on the same model class and loss scale. The upper expression is the selector rate, while the lower expression adds the exact-homogeneity baseline to the radius-sensitive rare-cell component. This form is useful for comparisons that keep d below the displayed calibrated range.

The endpoint reductions identify the two boundary cases of the finite-range bracket. Exact homogeneity reduces the selector and converse expressions to the collision rate, while radius two aligns the radius-indexed class with the unrestricted-radius class.

Proposition 4. [`prop:endpoint-reductions`] (Endpoint rate reductions) *For every overlap parameter $\epsilon \in (0, 1/2)$, there exist constants $c_\epsilon, c, C \in \mathbb{R}$ such that*

$$0 < c_\epsilon \le 1, \qquad 0 < c \le C.$$

Write

$$L_n = \log(en), \qquad u_{n,d} = \frac{d^2}{n^2 L_n^2}.$$

Then the following statements hold.

- ***(Exact-homogeneity endpoint.)*** *For all integers $n, d \ge 1$ such that*

  $$d \le c_\epsilon \frac{n^2}{L_n},$$

  the upper-frontier expression

  $$\frac{1}{n} + \min\left\{1, \min\left\{u_{n,d}, \frac{d}{n^2}\right\}\right\}$$

  and the capped converse expression

  $$\frac{1}{n} + \min\left\{1, \frac{d}{n^2}\right\}$$

  are both comparable to $n^{-1} + d/n^2$:

  $$c\left(\frac{1}{n} + \frac{d}{n^2}\right) \le \frac{1}{n} + \min\left\{1, \min\left\{u_{n,d}, \frac{d}{n^2}\right\}\right\} \le C\left(\frac{1}{n} + \frac{d}{n^2}\right),$$

  and

  $$c\left(\frac{1}{n} + \frac{d}{n^2}\right) \le \frac{1}{n} + \min\left\{1, \frac{d}{n^2}\right\} \le C\left(\frac{1}{n} + \frac{d}{n^2}\right).$$

- ***(Radius-two class identity.)*** *For every integer d and every $M \in \mathbb{R}$, the radius-two class in Definition 1 and the unrestricted class in Definition 2 have exactly the same attainable laws: every law in $\mathcal{P}^{\mathrm{unr}}_{d,\epsilon,M}$ is the law of some member of $\mathcal{P}_{d,\epsilon,M,2}$, and every law in $\mathcal{P}_{d,\epsilon,M,2}$ is the law of some member of $\mathcal{P}^{\mathrm{unr}}_{d,\epsilon,M}$.*

- ***(Radius-two endpoint.)*** *For all integers $n, d \geq 1$, the upper-frontier expression at radius two is exactly*

$$\frac{1}{n} + \min\{1, u_{n,d}\},$$

and the capped converse expression at radius two,

$$\frac{1}{n} + \min\left\{1, \frac{d}{n^2}\right\} + 4\min\{1, u_{n,d}\},$$

is comparable to the same rate:

$$c\left(\frac{1}{n} + \min\{1, u_{n,d}\}\right) \leq \frac{1}{n} + \min\left\{1, \frac{d}{n^2}\right\} + 4\min\{1, u_{n,d}\} \leq C\left(\frac{1}{n} + \min\{1, u_{n,d}\}\right).$$

Proposition 4 gives the finite-range endpoint algebra behind the rate discussion. At $\sigma = 0$, the polynomial branch and the collision lower benchmark reduce to $n^{-1} + d/n^2$ within the calibrated range. At $\sigma = 2$, the radius-indexed law class equals the unrestricted-radius class from Definition 2, and the upper expression is the polynomial large-alphabet rate.

The final result in this appendix localizes the regimes in which the finite-range bracket is order matched. It uses the baseline $b_{n,d}$, the polynomial scale $u_{n,d}$, the triangular selector benchmark $q_{n,d,\sigma}$, and the product-form triangular benchmark $\ell^{\mathrm{tri}}_{n,d,\sigma}$. The theorem statement below introduces the local shorthand $\ell_{n,d,\sigma}$ for that product-form benchmark.

Theorem 10. [`thm:fixed-interior-tightness-and-shrinking-radius-gap`] (Fixed-interior gap characterization) *For every overlap parameter ϵ with $0 < \epsilon < 1/2$, there is a constant $c_\epsilon > 0$ such that the following statements hold. Write*

$$L_n := \log(en), \qquad b_{n,d} := \frac{1}{n} + \frac{d}{n^2}, \qquad u_{n,d} := \frac{d^2}{n^2 L_n^2},$$

and

$$q_{n,d,\sigma} := \frac{1}{n} + \min\left\{u_{n,d},\ \sigma^2 + \frac{d}{n^2}\right\}, \qquad \ell_{n,d,\sigma} := b_{n,d} + \sigma^2 u_{n,d}.$$

Also let $r_{n,d,\sigma}$ denote the frontier rate

$$r_{n,d,\sigma} := \frac{1}{n} + \min\left\{1,\ \min\{u_{n,d},\ \left(\sigma^2 + \frac{d}{n^2}\right)\}\right\},$$

and let $\mathsf{R}_{n,d,\epsilon,M,\sigma}$ be the minimax mean-squared risk in Definition 5.

- ***(Triangular benchmark algebra.)*** *For all $n, d \in \mathbb{N}$ and all $\sigma \in [0, 2]$, if*

$$L_n^2 < d < nL_n,$$

then

$$\frac{1}{2}\left\{b_{n,d} + \min(u_{n,d}, \sigma^2)\right\} \leq q_{n,d,\sigma} \leq b_{n,d} + \min(u_{n,d}, \sigma^2), \qquad \ell_{n,d,\sigma} = b_{n,d} + \sigma^2 u_{n,d}.$$

- ***(Fixed interior radii.)*** *For every $\underline{\sigma} \in (0, 2]$, there are constants c, C with $0 < c \leq C$ such that, for all $n, d \in \mathbb{N}$ and $M, \sigma \in \mathbb{R}$, if*

$$n > 0, \qquad d > 0, \qquad M \geq 1, \qquad \underline{\sigma} \leq \sigma \leq 2, \qquad d \leq c_\epsilon \frac{n^2}{L_n},$$

then

$$cM^2 r_{n,d,\sigma} \leq \mathsf{R}_{n,d,\epsilon,M,\sigma} \leq CM^2 r_{n,d,\sigma}.$$

- ***(Matching elbows.)*** *For every $K \geq 0$, there are constants c, C with $0 < c \leq C$ such that, for all $n, d \in \mathbb{N}$ and $M, \sigma \in \mathbb{R}$, if*

$$n > 0, \qquad d > 0, \qquad M \geq 1, \qquad 0 \leq \sigma \leq 2,$$

and at least one of

$$u_{n,d} \geq 1, \qquad u_{n,d} \leq K b_{n,d}, \qquad \sigma^2 \leq K b_{n,d}$$

holds, then

$$cM^2 r_{n,d,\sigma} \leq \mathsf{R}_{n,d,\epsilon,M,\sigma} \leq CM^2 r_{n,d,\sigma}.$$

- ***(Exact algebraic elbow.)*** *For all $n, d \in \mathbb{N}$ with $n > 0$,*

$$u_{n,d} \leq b_{n,d} \quad \Longleftrightarrow \quad d^2 \leq nL_n^2 + dL_n^2.$$

- ***(Residual wedge implication.)*** *For every alphabet sequence d_t and radius sequence σ_t, if eventually*

$$d_t \geq 1, \qquad 0 \leq \sigma_t \leq 2,$$

and

$$\frac{r_{t,d_t,\sigma_t}}{t^{-1} + \min\{1, d_t/t^2\} + \sigma_t^2 \min\{1, u_{t,d_t}\}} \longrightarrow \infty,$$

then the sequence lies in the residual wedge:

$$\frac{b_{t,d_t}}{u_{t,d_t}} \longrightarrow 0, \qquad u_{t,d_t} \longrightarrow 0, \qquad \frac{b_{t,d_t}}{\sigma_t^2} \longrightarrow 0, \qquad \sigma_t^2 \longrightarrow 0.$$

- ***(Witness sequence.)*** *The sequence $d_n = n$ and $\sigma_n = L_n^{-1/2}$ lies in the residual wedge. Along this sequence,*

$$\frac{q_{n,n,L_n^{-1/2}}}{\ell_{n,n,L_n^{-1/2}}} \asymp L_n, \qquad \frac{q_{n,n,L_n^{-1/2}}}{\ell_{n,n,L_n^{-1/2}}} \longrightarrow \infty.$$

Theorem 10 gives the finite-range phase algebra supporting the matched-regime discussion in the main text. For every radius bounded below by a positive constant, the minimax risk is comparable to $M^2 r_{n,d,\sigma}$ over the calibrated alphabet range. The same comparison holds at the saturation and parametric-dominance elbows. In the triangular region $L_n^2 < d < nL_n$, the theorem separates the selector expression $b_{n,d} + \min(u_{n,d}, \sigma^2)$ from the converse expression $b_{n,d} + \sigma^2 u_{n,d}$, and the witness sequence shows that the residual shrinking-radius region has logarithmic benchmark separation.

D Verification note

This appendix records the scope of the machine-checked mathematical layer supporting the paper. The formalization covers the finite-cell observed-data model, the radius-indexed real-outcome classes, the estimators, the embeddings used for lower bounds, the selector risk bound, the endpoint reductions, the all-alphabet minimax bracket, and the regime-localization algebra. The displayed binary source lemmas are checked declarations in the current artifact; Zeng et al. [2024] supplies their statistical origin and attribution. Other citations supply source motivation and statistical context.

The setup objects are represented by Assumptions 1 to 6, together with the model, experiment, target, and risk definitions in Definitions 1 to 5. The scale-normalization step is represented by Lemma 1. These statements fix the statistical experiment whose minimax risk is studied throughout the paper.

The constructive side is represented by the estimator definitions in Algorithms 1 to 3 and by the risk guarantees in Theorems 1 and 2. The lower-bound side is represented by the embedding and least-favorable constructions in Algorithm 4 and Definitions 8 and 9, the binary-to-real transfer in Proposition 1, the exact-homogeneity lower transfer in Lemma 2, and the radius-channel converse in Theorem 3. The binary source lemmas in Lemmas 3 to 6 are checked declarations with faithful proof audits in the verification contract, while Zeng et al. [2024] supplies their origin and statistical context.

The paper's main synthesis is represented by Theorems 4 and 5 and Proposition 2. These statements assemble the all-alphabet bracket, identify the endpoint reductions, and characterize the fixed-radius, saturation, parametric-dominance, and residual shrinking-radius regimes. The binary source lemmas displayed above are checked declarations in this artifact, with citations supplying attribution and statistical context. The separate comparison proposition in Proposition 3 is checked as a conditional implication from the published binary collision guarantee displayed in its statement. The open regime description in Remark 1 is part of the recorded statement layer.

The anchored theorem, proposition, lemma, definition, and assumption statements displayed in the paper are machine-checked in Lean 4 at verification contract commit `462a124d29ae27887f711787ac2bc60f6b7e6e2f`. The main bracket in Theorem 4, selector upper bound in Theorem 2, radius-channel converse in Theorem 3, endpoint reductions in Proposition 2, and regime theorem in Theorem 5 are checked statements with faithful proof audits. The binary source lemmas in Lemmas 3 to 6 are discharged checked declarations in the artifact. The comparison proposition in Proposition 3 is checked as a conditional implication from its displayed published-collision hypothesis. Citations otherwise provide attribution and statistical context. The Lean toolchain is pinned by `lean-toolchain` to `leanprover/lean4:v4.33.0`; the repository build is invoked with `lake -d CausalSmith build`. The declaration names, source files, and paper object labels are recorded in `verification_contract.json` and `presentation_crosswalk.json` in the bundle.

E Proofs of the main results

Proof of Lemma 1. Fix $P \in \mathcal{P}^{\mathrm{unr}}_{d,\epsilon,M}$. Since $M \geq 1$, we have $M \geq 0$. For every supported cell k,

$$|\tau_k(P)| = |\mu_{1k}(P) - \mu_{0k}(P)| \leq |\mu_{1k}(P)| + |\mu_{0k}(P)| \leq M/2 + M/2 = M.$$

For every cell k, the product term in the all-cell formula of Definition 4 satisfies

$$|p_k(P)\tau_k(P)| \le p_k(P)M.$$

Indeed, the cell-mass range gives either $p_k(P) = 0$ or $p_k(P) > 0$. In the first case both sides are zero. In the second case,

$$|p_k(P)\tau_k(P)| = p_k(P)|\tau_k(P)| \le p_k(P)M,$$

using the supported-cell bound above and $p_k(P) > 0$. The finite cell masses satisfy

$$\sum_{k\in[d]} p_k(P) \le 1.$$

Hence, using the all-cell formula of Definition 4,

$$|\tau(P)| = \left|\sum_{k\in[d]} p_k(P)\tau_k(P)\right| \le \sum_{k\in[d]} |p_k(P)\tau_k(P)| \le \sum_{k\in[d]} p_k(P)M = \left(\sum_{k\in[d]} p_k(P)\right) M \le M.$$

Therefore, for every supported cell k,

$$|\delta_k(P)| = |\tau_k(P) - \tau(P)| \le |\tau_k(P)| + |\tau(P)| \le M + M = 2M.$$

The radius-two membership follows by keeping the same law and the same consistency, exchangeability, overlap, mean-normalization, and second-moment witnesses, using the numerical facts $0 \le 2 \le 2$, and adding the just-proved supported-cell bound

$$|\delta_k(P)| \le 2M = 2 \cdot M \qquad (p_k(P) > 0)$$

as the approximate-homogeneity witness. Thus every law in $\mathcal{P}^{\mathrm{unr}}_{d,\epsilon,M}$ is represented in $\mathcal{P}_{d,\epsilon,M,2}$. Conversely, a member of $\mathcal{P}_{d,\epsilon,M,2}$ carries all fields required for $\mathcal{P}^{\mathrm{unr}}_{d,\epsilon,M}$; forgetting the homogeneity and radius fields leaves the same law. $\square$

Proof of Theorem 7. 1. Fix an overlap parameter ϵ with $0 < \epsilon < 1/2$. By Theorem 1, choose $C^0_\epsilon > 0$ and a polynomial handle $(N_\epsilon, \rho_\epsilon)$, with $\rho_\epsilon > 0$, such that

$$N_\epsilon \le n, \qquad d \le \rho_\epsilon n \log(en) \quad\Longrightarrow\quad 2 \le K_n$$

for all positive integers n, d. The same construction gives, for every $n, d \ge 1$, every $M \ge 1$, and every $0 \le \sigma \le 2$, measurable $[-M, M]$-valued estimators T^{poly}_n and T^{col}_n, and for every $P \in \mathcal{P}_{d,\epsilon,M,\sigma}$,

$$E_{Q_P^{(n)}}\left[\left(T^{\mathrm{poly}}_n - \tau(P)\right)^2\right] \le C^0_\epsilon M^2 \left\{\frac{1}{n} + \min\left(1, \frac{d^2}{n^2\log^2(en)}\right)\right\}$$

and

$$E_{Q_P^{(n)}}\left[\left(T^{\mathrm{col}}_n - \tau(P)\right)^2\right] \le C^0_\epsilon M^2 \left\{\frac{1}{n} + \sigma^2 + \frac{d}{n^2}\right\}.$$

Set

$$C_\epsilon := \max\{1, C^0_\epsilon\}.$$

Then $C_\epsilon > 0$ and $C_\epsilon \ge C^0_\epsilon$, so both displayed risk bounds hold with C_ϵ in place of C^0_ϵ.

2. For fixed admissible n, d, M, σ, put

$$u_{n,d} := \frac{d^2}{n^2 \log^2(en)}, \qquad h_{n,d,\sigma} := \sigma^2 + \frac{d}{n^2}, \qquad r_{n,d,\sigma} := \frac{1}{n} + \min\{1, \min\{u_{n,d}, h_{n,d,\sigma}\}\}.$$

The known-radius selector of Algorithm 3 is the three-branch estimator

$$\widehat{\tau}_n^\star = \begin{cases} T_n^{\mathrm{col}}, & h_{n,d,\sigma} \le \min\{1, u_{n,d}\}, \\ T_n^{\mathrm{poly}}, & h_{n,d,\sigma} > \min\{1, u_{n,d}\} \text{ and } u_{n,d} < \min\{1, h_{n,d,\sigma}\}, \\ 0, & h_{n,d,\sigma} > \min\{1, u_{n,d}\} \text{ and } u_{n,d} \ge \min\{1, h_{n,d,\sigma}\}. \end{cases}$$

The first two branches are elements of $\mathcal{T}_{n,M}$ by the construction above. The zero branch is measurable and takes values in $[-M, M]$ because $M \ge 1$. Hence $\widehat{\tau}_n^\star \in \mathcal{T}_{n,M}$.

3. Fix $P \in \mathcal{P}_{d,\epsilon,M,\sigma}$. On the collision branch, $h_{n,d,\sigma} \le \min\{1, u_{n,d}\}$, so

$$\min\{1, \min\{u_{n,d}, h_{n,d,\sigma}\}\} = h_{n,d,\sigma}.$$

Therefore
$$E_{Q_P^{(n)}}\left[(\widehat{\tau}_n^\star - \tau(P))^2\right] \le C_\epsilon M^2 \left(\frac{1}{n} + h_{n,d,\sigma}\right) = C_\epsilon M^2 r_{n,d,\sigma}.$$

On the polynomial branch, $u_{n,d} < \min\{1, h_{n,d,\sigma}\}$, so

$$\min\{1, u_{n,d}\} = u_{n,d}, \qquad \min\{1, \min\{u_{n,d}, h_{n,d,\sigma}\}\} = u_{n,d}.$$

Thus
$$E_{Q_P^{(n)}}\left[(\widehat{\tau}_n^\star - \tau(P))^2\right] \le C_\epsilon M^2 \left(\frac{1}{n} + u_{n,d}\right) = C_\epsilon M^2 r_{n,d,\sigma}.$$

It remains to consider the zero branch. There

$$h_{n,d,\sigma} > \min\{1, u_{n,d}\}, \qquad u_{n,d} \ge \min\{1, h_{n,d,\sigma}\}.$$

These inequalities imply $1 \le u_{n,d}$: if $u_{n,d} < 1$, then $h_{n,d,\sigma} > u_{n,d}$, while $\min\{1, h_{n,d,\sigma}\} \le u_{n,d}$; if $h_{n,d,\sigma} \le 1$ this gives $h_{n,d,\sigma} \le u_{n,d} < h_{n,d,\sigma}$, and if $h_{n,d,\sigma} > 1$ it gives $1 \le u_{n,d} < 1$. Hence $1 \le u_{n,d}$. The first zero-branch inequality then gives $1 < h_{n,d,\sigma}$, so $1 \le h_{n,d,\sigma}$, and

$$\min\{1, \min\{u_{n,d}, h_{n,d,\sigma}\}\} = 1.$$

Viewing the same law as a member of the unrestricted-radius class, the scale bound in Lemma 1 gives
$$|\tau(P)| \le M.$$

Since the selected estimator is identically zero on this branch,

$$E_{Q_P^{(n)}}\left[(\widehat{\tau}_n^\star - \tau(P))^2\right] = \tau(P)^2 \le M^2 \le C_\epsilon M^2 \left(\frac{1}{n} + 1\right) = C_\epsilon M^2 r_{n,d,\sigma},$$

where the penultimate inequality uses $C_\epsilon \ge 1$ and $n \ge 1$. Together the three branches prove the asserted risk bound for every $P \in \mathcal{P}_{d,\epsilon,M,\sigma}$.

4. Put $c_\epsilon := 1$. Then $c_\epsilon > 0$. The preceding all-alphabet bound applies to every positive n, d, so it applies in particular under the restricted-dimensional hypothesis

$$d \le c_\epsilon \frac{n^2}{\log(en)}.$$

The constants C_ϵ, c_ϵ and the same polynomial handle $(N_\epsilon, \rho_\epsilon)$ therefore satisfy all claims of the theorem.

□

Proof of Proposition 4. 1. First record the logarithmic estimate used below. If $n \ge 1$, then

$$L_n = \log(en) = 1 + \log n.$$

The standard bound $\log n \le n^{1/2}/(1/2) = 2\sqrt{n}$, together with $\log n \ge 0$ and $\sqrt{n} \ge 1$, gives

$$0 \le L_n \le 1 + 2\sqrt{n} \le 3\sqrt{n}.$$

Squaring yields

$$L_n^2 \le 9n.$$

2. We prove the all-alphabet comparison that supplies the constants. For $n, d \ge 1$, put

$$\eta_{n,d} := \frac{d}{n^2}, \qquad u_{n,d} := \frac{d^2}{n^2 L_n^2}.$$

Since $L_n \ge 1$, both $\eta_{n,d}$ and $u_{n,d}$ are nonnegative. At radius two the collision term is

$$2^2 + \eta_{n,d} = 4 + \eta_{n,d} \ge 1,$$

and therefore

$$\frac{1}{n} + \min\{1, \min\{u_{n,d}, 4 + \eta_{n,d}\}\} = \frac{1}{n} + \min\{1, u_{n,d}\}.$$

At radius zero,

$$\frac{1}{n} + \min\{1, \min\{u_{n,d}, \eta_{n,d}\}\} \le \frac{1}{n} + \min\{1, \eta_{n,d}\},$$

and the capped converse expression at radius zero is exactly the right-hand side.

It remains to compare the radius-two converse term

$$\frac{1}{n} + \min\{1, \eta_{n,d}\} + 4\min\{1, u_{n,d}\}$$

with

$$\frac{1}{n} + \min\{1, u_{n,d}\}.$$

If $L_n^2 \le d$, then

$$\eta_{n,d} = \frac{d}{n^2} \le \frac{d^2}{n^2 L_n^2} = u_{n,d},$$

so

$$\min\{1, \min\{u_{n,d}, \eta_{n,d}\}\} = \min\{1, \eta_{n,d}\}$$

and

$$\min\{1, \eta_{n,d}\} \le \min\{1, u_{n,d}\}.$$

Consequently

$$\frac{1}{n} + \min\{1, u_{n,d}\} \le \frac{1}{n} + \min\{1, \eta_{n,d}\} + 4\min\{1, u_{n,d}\} \le 5\left(\frac{1}{n} + \min\{1, u_{n,d}\}\right).$$

If $d < L_n^2$, the logarithmic estimate from the previous step gives

$$d \le 9n, \qquad \eta_{n,d} \le \frac{9}{n}.$$

Hence

$$\frac{1}{n} + \min\{1, \eta_{n,d}\} \le \frac{10}{n}.$$

The upper-frontier term at radius zero is at least $1/n$, so

$$\frac{1}{10}\left(\frac{1}{n} + \min\{1, \eta_{n,d}\}\right) \le \frac{1}{n} + \min\{1, \min\{u_{n,d}, \eta_{n,d}\}\}.$$

The remaining upper bound at radius zero follows from the displayed monotonicity inequality. For radius two, nonnegativity gives the lower bound

$$\frac{1}{n} + \min\{1, u_{n,d}\} \le \frac{1}{n} + \min\{1, \eta_{n,d}\} + 4\min\{1, u_{n,d}\},$$

while $\min\{1, \eta_{n,d}\} \le 9/n$ gives

$$\frac{1}{n} + \min\{1, \eta_{n,d}\} + 4\min\{1, u_{n,d}\} \le 10\left(\frac{1}{n} + \min\{1, u_{n,d}\}\right).$$

Thus, uniformly over all $n, d \ge 1$, the all-alphabet comparisons hold with

$$c = \frac{1}{10}, \qquad C = 10.$$

The radius-two equality of attainable laws is exactly the consequence of Lemma 1.

3. Now fix $\epsilon \in (0, 1/2)$ and choose

$$c_\epsilon = 1, \qquad c = \frac{1}{10}, \qquad C = 10.$$

Then $0 < c_\epsilon \le 1$ and $0 < c \le C$. Suppose $n, d \ge 1$ and

$$d \le c_\epsilon \frac{n^2}{L_n} = \frac{n^2}{L_n}.$$

Because $L_n \ge 1$,

$$d \le n^2, \qquad \frac{d}{n^2} \le 1, \qquad \min\left\{1, \frac{d}{n^2}\right\} = \frac{d}{n^2}.$$

Applying the all-alphabet comparison from the previous step and replacing the capped term by d/n^2 gives

$$c\left(\frac{1}{n}+\frac{d}{n^2}\right) \le \frac{1}{n}+\min\left\{1, \min\left\{u_{n,d}, \frac{d}{n^2}\right\}\right\} \le C\left(\frac{1}{n}+\frac{d}{n^2}\right),$$

and

$$c\left(\frac{1}{n}+\frac{d}{n^2}\right) \le \frac{1}{n}+\min\left\{1, \frac{d}{n^2}\right\} \le C\left(\frac{1}{n}+\frac{d}{n^2}\right).$$

The class identity is the radius-two conclusion of Lemma 1. Finally, the all-alphabet radius-two calculation gives, for every $n, d \ge 1$,

$$\frac{1}{n}+\min\{1, \min\{u_{n,d}, 4+d/n^2\}\} = \frac{1}{n}+\min\{1, u_{n,d}\},$$

and its converse comparison gives

$$c\left(\frac{1}{n}+\min\{1, u_{n,d}\}\right) \le \frac{1}{n}+\min\left\{1, \frac{d}{n^2}\right\}+4\min\{1, u_{n,d}\} \le C\left(\frac{1}{n}+\min\{1, u_{n,d}\}\right).$$

$\square$

Proof of Theorem 8. 1. Fix an overlap parameter ϵ with $0<\epsilon<1/2$. Apply Theorem 3 to obtain constants $\bar{c}_\epsilon$ and $\bar{b}_\epsilon^{\mathrm{rad}}$ such that

$$0<\bar{c}_\epsilon \le 1, \qquad 0<\bar{b}_\epsilon^{\mathrm{rad}},$$

and, for every admissible n, d, M, σ,

$$\bar{c}_\epsilon M^2\left\{\frac{1}{n}+\min\left(1, \frac{d}{n^2}\right)+\sigma^2\min\left(1, \frac{d^2}{n^2\log^2(en)}\right)\right\} \le \mathsf{R}_{n,d,\epsilon,M,\sigma}.$$

Set $c_\epsilon := \bar{c}_\epsilon$ and $b_\epsilon^{\mathrm{rad}} := \bar{b}_\epsilon^{\mathrm{rad}}$.

2. Now fix n, d, M, σ satisfying the hypotheses of the present statement, and put

$$L := \log(en).$$

Since $0<n$, we have $n \ge 1$, hence

$$L = \log(en) = 1+\log n \ge 1.$$

The restricted range and $c_\epsilon \le 1$ give

$$d \le c_\epsilon \frac{n^2}{L} \le c_\epsilon n^2 \le n^2.$$

Because $n>0$, division by n^2 yields

$$\frac{d}{n^2} \le 1, \qquad \text{so} \qquad \min\left(1, \frac{d}{n^2}\right) = \frac{d}{n^2}.$$

Substituting this identity into the all-alphabet bound displayed above gives

$$\mathsf{R}_{n,d,\epsilon,M,\sigma} \ge c_\epsilon M^2\left\{\frac{1}{n}+\frac{d}{n^2}+\sigma^2\min\left(1, \frac{d^2}{n^2\log^2(en)}\right)\right\}.$$

3. The radial certificate supplied by Theorem 3 uses the same constants c_ϵ and $b_\epsilon^{\mathrm{rad}}$. Unpacking the certificate, there are an integer cap d_{rad}, a binary control-zero source class $\mathcal{S}_{\mathrm{rad}}$, a padding map into the d-alphabet, an embedding Ψ that sends each source member into $\mathcal{P}_{d,\epsilon,M,\sigma}$, and binary full-observation couplings such that the channel acts by

$$B'(a) \mid B(a) \sim \mathrm{Bernoulli}\left\{\frac{1}{2} + \frac{\sigma}{2}\left(B(a) - \frac{1}{2}\right)\right\}, \qquad Y(a) := M\{B'(a) - 1/2\}.$$

The source target in this certificate is the total treated-arm weighted functional

$$\psi(P) := \sum_{k=1}^{d_{\mathrm{rad}}} p_k(P)\, \mu_{1k}^{\mathrm{bin}}(P),$$

where $\mu_{1k}^{\mathrm{bin}}(P)$ is the treated-arm binary mean coordinate carried by the source law in cell k, defined for every source cell including zero-mass cells. Thus, for all $P_0, P_1 \in \mathcal{S}_{\mathrm{rad}}$, the same certificate gives

$$\tau(\Psi(P_1)) - \tau(\Psi(P_0)) = M\frac{\sigma}{2}\left\{\psi(P_1) - \psi(P_0)\right\}.$$

It also gives the estimator-wise lower bound: for every $\widehat{\tau} \in \mathcal{T}_{n,M}$, the class of total measurable $[-M, M]$-valued estimators from Definition 5, there is $P \in \mathcal{S}_{\mathrm{rad}}$ such that

$$c_\epsilon M^2 \sigma^2 \min\left(1, \frac{d^2}{n^2 \log^2(en)}\right) \le \mathbb{E}_{\Psi(P)}\left[\{\widehat{\tau} - \tau(\Psi(P))\}^2\right].$$

□

Proof of Theorem 6. Fix an overlap parameter ϵ with $0 < \epsilon < 1/2$.

1. First form the all-alphabet package. By Lemma 13, there are $C_{\mathrm{poly},\epsilon} > 0$ and a polynomial handle

$$\mathfrak{h}_\epsilon = (N_{\mathrm{cal},\epsilon}, \rho_{\mathrm{cal},\epsilon}), \qquad \rho_{\mathrm{cal},\epsilon} > 0,$$

such that, for all integers $n, d \ge 1$,

$$N_{\mathrm{cal},\epsilon} \le n, \qquad d \le \rho_{\mathrm{cal},\epsilon} n \log(en) \quad \Longrightarrow \quad 2 \le K_n,$$

and, for every $M \ge 1$, $0 \le \sigma \le 2$, and $P \in \mathcal{P}_{d,\epsilon,M,\sigma}$,

$$E_{Q_P^{(n)}}\left[\left(T_n^{\mathrm{poly}} - \tau(P)\right)^2\right] \le C_{\mathrm{poly},\epsilon} M^2 \left\{\frac{1}{n} + \min\left(1, \frac{d^2}{n^2 \log^2(en)}\right)\right\}.$$

The same cited construction gives measurability of T_n^{poly} and

$$T_n^{\mathrm{poly}}(s) \in [-M, M] \qquad \text{for every } s \in \mathcal{O}_d^n.$$

By Lemma 12, there is $C_{\mathrm{col},\epsilon} > 0$ such that, for every $n, d \ge 1$, $M \ge 1$, $0 \le \sigma \le 2$, and $P \in \mathcal{P}_{d,\epsilon,M,\sigma}$,

$$E_{Q_P^{(n)}}\left[\left(T_n^{\mathrm{col}} - \tau(P)\right)^2\right] \le C_{\mathrm{col},\epsilon} M^2 \left(\frac{1}{n} + \sigma^2 + \frac{d}{n^2}\right).$$

For the admissibility of the collision estimator, write its unclipped inner statistic as the total sample function

$$\Gamma_{\mathrm{col}}(s) := \begin{cases} \dfrac{1}{R_n(s)} \displaystyle\sum_{k=1}^{d} N_k(s) I_k(s) \left[\mathbf{1}\{N_{1k}(s) > 0\} \frac{S_{1k}(s)}{N_{1k}(s) \vee 1} - \mathbf{1}\{N_{0k}(s) > 0\} \frac{S_{0k}(s)}{N_{0k}(s) \vee 1} \right], & R_n(s) > 0, \\ 0, & R_n(s) = 0. \end{cases}$$

Here $m \vee 1$ denotes $\max\{m, 1\}$. The displayed arm factors are total functions of the sample and agree with the arm-cell means of Algorithm 2 on positive arm-cell counts; multiplication by I_k selects exactly the cells where both arm-cell counts are positive. Hence Algorithm 2 gives

$$T_n^{\mathrm{col}}(s) = \max\{-M, \min(M, \Gamma_{\mathrm{col}}(s))\}.$$

The counts, indicators, outcome totals, finite sums, totalized ratios, and the fallback at $R_n = 0$ are measurable functions of the sample, so T_n^{col} is measurable. Since $M \geq 1$, clipping gives

$$T_n^{\mathrm{col}}(s) \in [-M, M] \qquad \text{for every } s \in \mathcal{O}_d^n.$$

Set

$$C_\epsilon := \max\{C_{\mathrm{poly},\epsilon}, C_{\mathrm{col},\epsilon}\}.$$

Then $C_\epsilon > 0$. The two displayed rate factors are nonnegative for $n, d \geq 1$ and $0 \leq \sigma \leq 2$, and $M^2 \geq 0$; hence replacing $C_{\mathrm{poly},\epsilon}$ and $C_{\mathrm{col},\epsilon}$ by C_ϵ preserves the two upper bounds.

2. Define

$$c_\epsilon := 1.$$

Then $c_\epsilon > 0$. For arbitrary $n, d \geq 1$, $M \geq 1$, and $0 \leq \sigma \leq 2$, the all-alphabet package from the first step gives the measurability and range assertions for both estimators. If

$$d \leq c_\epsilon \frac{n^2}{\log(en)},$$

then for every $P \in \mathcal{P}_{d,\epsilon,M,\sigma}$ the same all-alphabet risk bounds apply under this range condition:

$$E_{Q_P^{(n)}}\left[\left(T_n^{\mathrm{poly}} - \tau(P)\right)^2\right] \leq C_\epsilon M^2 \left\{\frac{1}{n} + \min\left(1, \frac{d^2}{n^2 \log^2(en)}\right)\right\},$$

and

$$E_{Q_P^{(n)}}\left[\left(T_n^{\mathrm{col}} - \tau(P)\right)^2\right] \leq C_\epsilon M^2 \left(\frac{1}{n} + \sigma^2 + \frac{d}{n^2}\right).$$

Together with the active-range property of $\mathfrak{h}_\epsilon$, these are the asserted constants and calibrated handle.

□

Proof of Theorem 10. 1. Fix $0 < \epsilon < 1/2$, and apply Theorem 5 at this ϵ. Choose the range constant

$$c_\epsilon := 1.$$

Then $c_\epsilon > 0$. The triangular benchmark algebra, the exact algebraic elbow, the residual wedge implication, and the diagonal witness assertions in the present statement are exactly the corresponding conclusions supplied by Theorem 5, with the same definitions of L_n, $b_{n,d}$, $u_{n,d}$, $q_{n,d,\sigma}$, $\ell_{n,d,\sigma}$, and $r_{n,d,\sigma}$. In particular, the denominator in the residual-wedge ratio is the capped converse benchmark

$$\frac{1}{t} + \min\left\{1, \frac{d_t}{t^2}\right\} + \sigma_t^2 \min\{1, u_{t,d_t}\},$$

so the all-alphabet residual-wedge implication gives precisely

$$\frac{b_{t,d_t}}{u_{t,d_t}} \to 0, \qquad u_{t,d_t} \to 0, \qquad \frac{b_{t,d_t}}{\sigma_t^2} \to 0, \qquad \sigma_t^2 \to 0.$$

For the witness $d_n = n$, $\sigma_n = L_n^{-1/2}$, the same all-alphabet conclusion gives membership in the residual wedge and the two displayed comparisons

$$\frac{q_{n,n,L_n^{-1/2}}}{\ell_{n,n,L_n^{-1/2}}} \asymp L_n, \qquad \frac{q_{n,n,L_n^{-1/2}}}{\ell_{n,n,L_n^{-1/2}}} \to \infty.$$

2. It remains only to specialize the fixed-interior and matching-elbow minimax clauses. Let $\underline{\sigma} \in (0, 2]$. The fixed-radius clause of Theorem 5 supplies constants c, C with $0 < c \le C$ such that, for all $n, d \in \mathbb{N}$, $M, \sigma \in \mathbb{R}$,

$$n > 0, \qquad d > 0, \qquad M \ge 1, \qquad 0 \le \sigma, \qquad \underline{\sigma} \le \sigma, \qquad \sigma \le 2$$

imply

$$cM^2 r_{n,d,\sigma} \le \mathsf{R}_{n,d,\epsilon,M,\sigma} \le CM^2 r_{n,d,\sigma}.$$

Since the implication holds for every positive d, it also holds under the additional displayed range condition

$$d \le c_\epsilon \frac{n^2}{L_n} = \frac{n^2}{L_n}.$$

Thus the fixed-interior-radii conclusion follows with the same constants c, C.

3. Now fix $K \ge 0$. The matching-elbow clause of Theorem 5 supplies constants c, C with $0 < c \le C$ such that, whenever

$$n > 0, \qquad d > 0, \qquad M \ge 1, \qquad 0 \le \sigma \le 2$$

and at least one of

$$u_{n,d} \ge 1, \qquad u_{n,d} \le K b_{n,d}, \qquad \sigma^2 \le K b_{n,d}$$

holds, one has

$$cM^2 r_{n,d,\sigma} \le \mathsf{R}_{n,d,\epsilon,M,\sigma} \le CM^2 r_{n,d,\sigma}.$$

This is exactly the matching-elbows assertion in the present theorem. Combining this with the clauses recorded in the first step and the choice $c_\epsilon = 1$ proves all asserted conclusions. □

Proof of Theorem 9. 1. Fix an overlap parameter ϵ with $0 < \epsilon < 1/2$. By Theorem 4, choose constants c_ϵ and C_ϵ such that

$$0 < c_\epsilon \le 1, \qquad 1 \le C_\epsilon, \qquad c_\epsilon \le C_\epsilon,$$

and such that, for every $n, d \ge 1$, $M \ge 1$, and $0 \le \sigma \le 2$,

$$c_\epsilon M^2 \left\{ \frac{1}{n} + \min\left(1, \frac{d}{n^2}\right) + \sigma^2 \min\left(1, \frac{d^2}{n^2 \log^2(en)}\right) \right\} \le \mathsf{R}_{n,d,\epsilon,M,\sigma}$$

and

$$\mathsf{R}_{n,d,\epsilon,M,\sigma} \le C_\epsilon M^2 \left\{ \frac{1}{n} + \min\left(1, \min\left(\frac{d^2}{n^2 \log^2(en)}, \sigma^2 + \frac{d}{n^2}\right)\right) \right\}.$$

These constants already have the normalization required in the statement.

2. Now fix $n, d \ge 1$, $M \ge 1$, and $0 \le \sigma \le 2$, and assume

$$d \le \frac{c_\epsilon n^2}{\log(en)}.$$

Since $n \ge 1$,

$$\log(en) = \log(e) + \log n = 1 + \log n \ge 1.$$

The numerator $c_\epsilon n^2$ is nonnegative, so

$$d \le \frac{c_\epsilon n^2}{\log(en)} \le c_\epsilon n^2 \le n^2,$$

where the last inequality uses $c_\epsilon \le 1$. Dividing by $n^2 > 0$ gives

$$\frac{d}{n^2} \le 1, \qquad \text{and hence} \qquad \min\left(1, \frac{d}{n^2}\right) = \frac{d}{n^2}.$$

Substituting this identity into the lower bound from the preceding step yields

$$c_\epsilon M^2 \left\{ \frac{1}{n} + \frac{d}{n^2} + \sigma^2 \min\left(1, \frac{d^2}{n^2 \log^2(en)}\right) \right\} \le \mathsf{R}_{n,d,\epsilon,M,\sigma}.$$

The upper bound is exactly the upper bound supplied in the preceding step. Therefore the displayed lower and upper bounds hold throughout the stated restricted range.

□

Proof of Lemma 2. Fix $0 < \epsilon < 1/2$. Let $a_\epsilon, b_\epsilon > 0$ and $N_\epsilon \in \mathbb{N}$ be the constants supplied by Lemma 4. Let $c_{\text{par}} > 0$ be the constant in Lemma 14. Set

$$\bar{b}_\epsilon := \min\{b_\epsilon, 1/2\}, \qquad K_\epsilon := 1 + N_\epsilon + \frac{1}{\bar{b}_\epsilon}, \qquad c_\epsilon := \min\left\{ \frac{a_\epsilon \bar{b}_\epsilon}{4}, \frac{c_{\text{par}}}{K_\epsilon} \right\}.$$

Then $\bar{b}_\epsilon > 0$, $K_\epsilon > 0$, and $c_\epsilon > 0$.

Fix $n, d \ge 1$, $M \ge 1$, and $0 \le \sigma \le 2$. Put

$$x := \bar{b}_\epsilon n^2.$$

First suppose that

$$N_\epsilon \leq n, \qquad 1 \leq x.$$

Define the capped source alphabet size

$$d_0 := \min\{d, \lfloor x \rfloor\}.$$

Then $d_0 \geq 1$, $d_0 \leq d$, and

$$d_0 \leq b_\epsilon n^2.$$

The exact binary source bound gives

$$a_\epsilon \left(\frac{1}{n} + \frac{d_0}{n^2} \right) \leq \mathsf{R}^{\mathrm{bin}}_{n,d_0,\epsilon},$$

where $\mathsf{R}^{\mathrm{bin}}_{n,d_0,\epsilon}$ denotes the minimax risk over the uniform-mass exactly homogeneous binary source experiment. Since $a_\epsilon > 0$ and $n, d_0 \geq 1$, the source lower-risk level

$$L_{\mathrm{ex}} := \frac{a_\epsilon}{2} \left(\frac{1}{n} + \frac{d_0}{n^2} \right)$$

is strictly below $\mathsf{R}^{\mathrm{bin}}_{n,d_0,\epsilon}$. Hence, for every measurable binary estimator there is a source law P such that

$$\frac{a_\epsilon}{2} \left(\frac{1}{n} + \frac{d_0}{n^2} \right) \leq E_P \left[\left\{ \widehat{\psi} - \psi(P) \right\}^2 \right].$$

Embed this source law into the d-cell real-outcome experiment by zero-mass padding and by the affine outcome map

$$Y(a) = M\{B(a) - 1/2\}, \qquad a \in \{0, 1\}.$$

The padding keeps the source cells inside the first d_0 target cells and assigns zero mass to cells $d_0+1, \ldots, d$. The affine map gives conditional means in $[-M/2, M/2]$, conditional second central moments bounded by M^2, and exact homogeneity; with the assumed bounds on ϵ, M, and σ, the embedded law belongs to

$$\mathcal{P}_{d,\epsilon,M,\sigma}.$$

Write $\Phi_M(P)$ for this embedded real law. Its target satisfies

$$\tau(\Phi_M(P)) = M\psi(P),$$

and the observed-data law is the pushforward of the binary observed-data law under the padded affine sample map. Thus every $T \in \mathcal{T}_{n,M}$ from Definition 5 induces a measurable binary estimator by applying $M^{-1}T$ after that sample map, and

$$M^2 \frac{a_\epsilon}{2} \left(\frac{1}{n} + \frac{d_0}{n^2} \right) \leq \mathsf{R}_{n,d,\epsilon,M,\sigma}.$$

It remains in this branch to compare the capped source rate with the desired all-d rate. Since $1 \leq x$,

$$\frac{x}{2} \leq \lfloor x \rfloor.$$

If $d \le \lfloor x \rfloor$, then $d_0 = d$ and $d/n^2 \le \bar{b}_\epsilon \le 1$. If $d > \lfloor x \rfloor$, then $d_0 = \lfloor x \rfloor$ and $d_0/n^2 \ge \bar{b}_\epsilon/2$. In both cases,

$$\frac{\bar{b}_\epsilon}{2}\left\{\frac{1}{n} + \min\left(1, \frac{d}{n^2}\right)\right\} \le \frac{1}{n} + \frac{d_0}{n^2}.$$

Since $c_\epsilon \le a_\epsilon \bar{b}_\epsilon/4$,

$$c_\epsilon M^2\left\{\frac{1}{n} + \min\left(1, \frac{d}{n^2}\right)\right\} \le M^2\frac{a_\epsilon}{2}\left(\frac{1}{n} + \frac{d_0}{n^2}\right) \le \mathsf{R}_{n,d,\epsilon,M,\sigma}.$$

On the complementary branch, either $n < N_\epsilon$ or $\bar{b}_\epsilon n^2 < 1$. In the first case, $N_\epsilon/n \ge 1$; in the second, $\bar{b}_\epsilon n \le 1$, so $(1/\bar{b}_\epsilon)/n \ge 1$. Thus, in either case,

$$\frac{1}{n} + \min\left(1, \frac{d}{n^2}\right) \le \frac{K_\epsilon}{n}.$$

Using $c_\epsilon \le c_{\mathrm{par}}/K_\epsilon$ and Lemma 14,

$$c_\epsilon M^2\left\{\frac{1}{n} + \min\left(1, \frac{d}{n^2}\right)\right\} \le \frac{c_{\mathrm{par}} M^2}{n} \le \mathsf{R}_{n,d,\epsilon,M,\sigma}.$$

Combining the two cases proves the stated bound for all $n, d \ge 1$, $M \ge 1$, and $0 \le \sigma \le 2$. □

Proof of Theorem 1. Fix an overlap parameter $0 < \epsilon < 1/2$.

1. By Lemma 13, choose $C_\epsilon^{\mathrm{poly}} > 0$ and a polynomial handle $h_\epsilon = (N_\epsilon, \rho_\epsilon)$. The handle has $\rho_\epsilon > 0$ and satisfies the calibration clause

   $$N_\epsilon \le n, \qquad d \le \rho_\epsilon n \log(en) \quad \Longrightarrow \quad 2 \le K_n$$

   for all integers $n, d \ge 1$. The same lemma gives, for every $n, d \ge 1$, $M \ge 1$, $0 \le \sigma \le 2$, and $P \in \mathcal{P}_{d,\epsilon,M,\sigma}$, that the estimator T_n^{poly} indexed by this handle is measurable, takes values in $[-M, M]$, and satisfies

   $$E_{Q_P^{(n)}}\left[\{T_n^{\mathrm{poly}} - \tau(P)\}^2\right] \le C_\epsilon^{\mathrm{poly}} M^2\left\{\frac{1}{n} + \min\left(1, \frac{d^2}{n^2\log^2(en)}\right)\right\}.$$

2. By Lemma 12, choose $C_\epsilon^{\mathrm{col}} > 0$ such that, for every $n, d \ge 1$ and every $P \in \mathcal{P}_{d,\epsilon,M,\sigma}$,

   $$E_{Q_P^{(n)}}\left[\{T_n^{\mathrm{col}} - \tau(P)\}^2\right] \le C_\epsilon^{\mathrm{col}} M^2\left(\frac{1}{n} + \sigma^2 + \frac{d}{n^2}\right).$$

   Here the bounds $M \ge 1$ and $0 \le \sigma \le 2$ are the corresponding model-class fields in $\mathcal{P}_{d,\epsilon,M,\sigma}$ from Definition 1.

3. The collision estimator also has the required admissibility property. For a sample s, Algorithm 2 defines

   $$T_n^{\mathrm{col}}(s) = \mathrm{clip}_{[-M,M]}\left(\begin{cases}\dfrac{\sum_{k=1}^d \mathbf{1}\{N_{0k}(s) > 0,\ N_{1k}(s) > 0\} N_k(s)\left(\overline{Y}_{1k}(s) - \overline{Y}_{0k}(s)\right)}{R_n(s)}, & R_n(s) > 0,\\[2ex] 0, & R_n(s) = 0.\end{cases}\right),$$

where each empirical arm-cell mean is the indicator-totalized ratio $S_{ak}(s)/N_{ak}(s)$ on $N_{ak}(s) > 0$ and is assigned value 0 when $N_{ak}(s) = 0$. The counts are finite sums of sample indicators, the totals are finite sums of indicator-weighted outcomes, and the branches use measurable comparisons, arithmetic operations, and a finite sum over cells. Thus the inner statistic is a measurable function of the n-sample. Since $M \geq 1$ gives $0 \leq M$, clipping to $[-M, M]$ is measurable and has range contained in $[-M, M]$. Therefore T_n^{col} is a measurable map of the sample and satisfies

$$T_n^{\mathrm{col}}(s) \in [-M, M] \qquad \text{for every sample } s.$$

4. Define

$$C_\epsilon := \max\{C_\epsilon^{\mathrm{poly}}, C_\epsilon^{\mathrm{col}}\}.$$

Then $C_\epsilon > 0$. For $n, d \geq 1$ and $0 \leq \sigma \leq 2$,

$$0 \leq \frac{1}{n} + \min\left(1, \frac{d^2}{n^2 \log^2(en)}\right), \qquad 0 \leq \frac{1}{n} + \sigma^2 + \frac{d}{n^2}.$$

Together with $M^2 \geq 0$, the inequalities $C_\epsilon^{\mathrm{poly}} \leq C_\epsilon$ and $C_\epsilon^{\mathrm{col}} \leq C_\epsilon$ upgrade the two pointwise risk inequalities above to the same inequalities with C_ϵ. With $h = h_\epsilon$, the calibration, measurability, range, and pointwise risk clauses hold simultaneously for every $n, d \geq 1$, $M \geq 1$, $0 \leq \sigma \leq 2$, and every $P \in \mathcal{P}_{d,\epsilon,M,\sigma}$ from Definition 1. Since the resulting right-hand sides do not depend on P, taking the supremum over $P \in \mathcal{P}_{d,\epsilon,M,\sigma}$ gives the two displayed supremum inequalities.

□

Proof of Theorem 2. Fix $0 < \epsilon < 1/2$.

Step 1. Common calibration and envelopes. By Theorem 1, choose a polynomial handle $(N_\epsilon, \rho_\epsilon)$, with $\rho_\epsilon > 0$, and a constant $\Gamma_\epsilon > 0$. Put

$$C_\epsilon := \max\{1, \Gamma_\epsilon\}.$$

Then $C_\epsilon > 0$, $\Gamma_\epsilon \leq C_\epsilon$, and $1 \leq C_\epsilon$. The selected handle carries the displayed calibration implication for K_n: for all integers $n, d \geq 1$,

$$N_\epsilon \leq n, \qquad d \leq \rho_\epsilon n \log(en) \quad \Longrightarrow \quad 2 \leq K_n.$$

For fixed integers $n, d \geq 1$, $M \geq 1$, and $0 \leq \sigma \leq 2$, define

$$u_{n,d} := \frac{d^2}{n^2 \log^2(en)}, \qquad h_{n,d,\sigma} := \sigma^2 + \frac{d}{n^2}.$$

The construction also gives measurable $[-M, M]$-valued estimators T_n^{poly} and T_n^{col}, and for every $P \in \mathcal{P}_{d,\epsilon,M,\sigma}$,

$$E_{Q_P^{(n)}}\left[\left(T_n^{\mathrm{poly}} - \tau(P)\right)^2\right] \leq \Gamma_\epsilon M^2 \left(\frac{1}{n} + \min\{1, u_{n,d}\}\right)$$

and

$$E_{Q_P^{(n)}}\left[\left(T_n^{\mathrm{col}} - \tau(P)\right)^2\right] \leq \Gamma_\epsilon M^2 \left(\frac{1}{n} + h_{n,d,\sigma}\right).$$

Step 2. The selector is admissible. The known-radius rule in Algorithm 3 is

$$\widehat{\tau}_n^\star = \begin{cases} T_n^{\mathrm{col}}, & h_{n,d,\sigma} \le \min\{1, u_{n,d}\}, \\ T_n^{\mathrm{poly}}, & u_{n,d} < \min\{1, h_{n,d,\sigma}\}, \\ 0, & \text{otherwise}. \end{cases}$$

The first two branches are measurable and take values in $[-M, M]$ by Theorem 1. The zero branch is measurable and also takes values in $[-M, M]$ because $M \ge 1$. Hence $\widehat{\tau}_n^\star$ is measurable and $\widehat{\tau}_n^\star(s) \in [-M, M]$ for every sample s.

Step 3. Collision branch. Assume

$$h_{n,d,\sigma} \le \min\{1, u_{n,d}\}.$$

Then $\widehat{\tau}_n^\star = T_n^{\mathrm{col}}$, and

$$\min\{1, \min(u_{n,d}, h_{n,d,\sigma})\} = h_{n,d,\sigma}.$$

Therefore, for every $P \in \mathcal{P}_{d,\epsilon,M,\sigma}$,

$$E_{Q_P^{(n)}}\left[(\widehat{\tau}_n^\star - \tau(P))^2\right] \le \Gamma_\epsilon M^2 \left(\frac{1}{n} + h_{n,d,\sigma}\right) \le C_\epsilon M^2 \left(\frac{1}{n} + \min\{1, \min(u_{n,d}, h_{n,d,\sigma})\}\right),$$

where the last inequality uses $\Gamma_\epsilon \le C_\epsilon$ and the nonnegativity of $M^2(1/n + h_{n,d,\sigma})$.

Step 4. Polynomial branch. Assume the collision-branch condition fails and

$$u_{n,d} < \min\{1, h_{n,d,\sigma}\}.$$

Then $\widehat{\tau}_n^\star = T_n^{\mathrm{poly}}$,

$$\min\{1, u_{n,d}\} = u_{n,d}, \qquad \min\{1, \min(u_{n,d}, h_{n,d,\sigma})\} = u_{n,d}.$$

Hence, for every $P \in \mathcal{P}_{d,\epsilon,M,\sigma}$,

$$E_{Q_P^{(n)}}\left[(\widehat{\tau}_n^\star - \tau(P))^2\right] \le \Gamma_\epsilon M^2 \left(\frac{1}{n} + u_{n,d}\right) \le C_\epsilon M^2 \left(\frac{1}{n} + \min\{1, \min(u_{n,d}, h_{n,d,\sigma})\}\right),$$

again because $\Gamma_\epsilon \le C_\epsilon$ and the rate factor is nonnegative.

Step 5. Zero branch. It remains to consider the branch in which

$$h_{n,d,\sigma} > \min\{1, u_{n,d}\}, \qquad u_{n,d} \ge \min\{1, h_{n,d,\sigma}\}.$$

These two inequalities imply $1 \le u_{n,d}$: if $u_{n,d} < 1$, then the first inequality gives $h_{n,d,\sigma} > u_{n,d}$, while the second gives $\min\{1, h_{n,d,\sigma}\} \le u_{n,d}$, a contradiction whether $h_{n,d,\sigma} \le 1$ or $h_{n,d,\sigma} > 1$. With $1 \le u_{n,d}$, the first displayed inequality gives $1 \le h_{n,d,\sigma}$, and consequently

$$\min\{1, \min(u_{n,d}, h_{n,d,\sigma})\} = 1.$$

On this branch $\widehat{\tau}_n^\star = 0$. Viewing the law of $P \in \mathcal{P}_{d,\epsilon,M,\sigma}$ as a member of the unrestricted class obtained by forgetting the radius condition, Lemma 1 gives

$$|\tau(P)| \le M.$$

Thus

$$E_{Q_P^{(n)}}\left[(\widehat{\tau}_n^\star - \tau(P))^2\right] = \tau(P)^2 \le M^2.$$

Since $C_\epsilon \ge 1$, $M^2 \ge 0$, and $n \ge 1$,

$$M^2 \le C_\epsilon M^2 \left(\frac{1}{n} + 1\right) = C_\epsilon M^2 \left(\frac{1}{n} + \min\{1, \min(u_{n,d}, h_{n,d,\sigma})\}\right).$$

Combining the three branches yields, for every $P \in \mathcal{P}_{d,\epsilon,M,\sigma}$,

$$E_{Q_P^{(n)}}\left[(\widehat{\tau}_n^\star - \tau(P))^2\right] \le C_\epsilon M^2 \left(\frac{1}{n} + \min\{1, \min(u_{n,d}, h_{n,d,\sigma})\}\right) = C_\epsilon M^2 r_{n,d,\sigma}.$$

Together with the measurability and range conclusion in Step 2, this proves the stated selector guarantee. □

Proof of Proposition 2. Take $c = 1/10$ and $C = 10$. Fix integers $n, d \ge 1$, and write

$$\Lambda_n := \log(en), \qquad u_{n,d} := \frac{d^2}{n^2\Lambda_n^2}, \qquad v_{n,d} := \frac{d}{n^2}.$$

Also set

$$\theta_{n,d}^{\mathrm{ex}} := \frac{1}{n} + \min(1, v_{n,d}), \qquad \theta_{n,d}^{\mathrm{poly}} := \frac{1}{n} + \min(1, u_{n,d}).$$

The quantities $1/n$, $u_{n,d}$, and $v_{n,d}$ are nonnegative.

1. Since $n \ge 1$,

   $$\Lambda_n = \log(en) = 1 + \log n \ge 1,$$

   and hence $\Lambda_n^2 > 0$. At radius two the collision term is

   $$2^2 + v_{n,d} = 4 + v_{n,d} \ge 1.$$

   Therefore

   $$\frac{1}{n} + \min\{1, u_{n,d}, 4 + v_{n,d}\} = \frac{1}{n} + \min(1, u_{n,d}) = \theta_{n,d}^{\mathrm{poly}}.$$

   This proves the endpoint identity, and also gives the two-sided comparison of the radius-two upper frontier expression with $\theta_{n,d}^{\mathrm{poly}}$, because $\theta_{n,d}^{\mathrm{poly}} \ge 0$ and $1/10 \le 1 \le 10$.

2. We use the elementary logarithmic estimate

   $$\Lambda_n^2 \le 9n.$$

   Indeed, $\log n \le 2\sqrt{n}$, while $\sqrt{n} \ge 1$ and $\log n \ge 0$. Thus

   $$\Lambda_n = 1 + \log n \le 1 + 2\sqrt{n} \le 3\sqrt{n},$$

   and squaring gives $\Lambda_n^2 \le 9n$.

3. First consider the case $\Lambda_n^2 \le d$. Since $d > 0$,

$$v_{n,d} = \frac{d}{n^2} \le \frac{d^2}{n^2\Lambda_n^2} = u_{n,d}.$$

Consequently

$$\frac{1}{n} + \min\{1, u_{n,d}, v_{n,d}\} = \frac{1}{n} + \min(1, v_{n,d}) = \theta^{\mathrm{ex}}_{n,d}.$$

The radius-zero converse expression is also exactly

$$\frac{1}{n} + \min(1, v_{n,d}) = \theta^{\mathrm{ex}}_{n,d}.$$

For the radius-two converse expression,

$$\frac{1}{n} + \min(1, v_{n,d}) + 4\min(1, u_{n,d}) \ge \theta^{\mathrm{poly}}_{n,d},$$

and, using $\min(1, v_{n,d}) \le \min(1, u_{n,d})$,

$$\frac{1}{n} + \min(1, v_{n,d}) + 4\min(1, u_{n,d}) \le \frac{1}{n} + 5\min(1, u_{n,d}) \le 5\theta^{\mathrm{poly}}_{n,d}.$$

All required comparisons in this case follow from $1/10 \le 1 \le 5 \le 10$.

4. Now consider the case $d < \Lambda_n^2$. By the logarithmic estimate,

$$d \le 9n, \qquad \text{so} \qquad v_{n,d} \le \frac{9}{n}.$$

The radius-zero upper expression satisfies

$$\frac{1}{n} \le \frac{1}{n} + \min\{1, u_{n,d}, v_{n,d}\} \le \frac{1}{n} + \min(1, v_{n,d}) = \theta^{\mathrm{ex}}_{n,d},$$

and

$$\theta^{\mathrm{ex}}_{n,d} = \frac{1}{n} + \min(1, v_{n,d}) \le \frac{1}{n} + \frac{9}{n} = \frac{10}{n}.$$

Hence

$$\frac{1}{10}\theta^{\mathrm{ex}}_{n,d} \le \frac{1}{n} \le \frac{1}{n} + \min\{1, u_{n,d}, v_{n,d}\} \le \theta^{\mathrm{ex}}_{n,d} \le 10\theta^{\mathrm{ex}}_{n,d}.$$

The radius-zero converse expression again equals $\theta^{\mathrm{ex}}_{n,d}$. For the radius-two converse expression,

$$\frac{1}{n} + \min(1, v_{n,d}) + 4\min(1, u_{n,d}) \ge \theta^{\mathrm{poly}}_{n,d},$$

and

$$\frac{1}{n} + \min(1, v_{n,d}) + 4\min(1, u_{n,d}) \le \frac{10}{n} + 4\min(1, u_{n,d}) \le 10\theta^{\mathrm{poly}}_{n,d}.$$

Together with the radius-two identity from the first step, this proves the stated two-sided comparisons in the second case.

5. It remains to identify the two radius-two classes. By Lemma 1, for every d, every $\epsilon \in \mathbb{R}$, and every $M \in \mathbb{R}$, each law in $\mathcal{P}^{\mathrm{unr}}_{d,\epsilon,M}$ is represented by some member of $\mathcal{P}_{d,\epsilon,M,2}$ with the same law, and conversely each law in $\mathcal{P}_{d,\epsilon,M,2}$ is represented by some member of $\mathcal{P}^{\mathrm{unr}}_{d,\epsilon,M}$ with the same law. Thus the represented laws are exactly the same at radius two.

$\square$

Proof of Theorem 5. Fix $0 < \epsilon < 1/2$. The proof starts from the all-alphabet bracket in Theorem 4; write

$$\omega_{n,d,\sigma} := \frac{1}{n} + \min\left(1, \frac{d}{n^2}\right) + \sigma^2 \min(1, u_{n,d})$$

for the capped converse benchmark appearing there.

In the triangular region $L_n^2 < d < nL_n$, set $a = 1/n$, $t = d/n^2$, $u = u_{n,d}$, and $s = \sigma^2$. The two triangular inequalities give $t \leq u < 1$, and the selector benchmark is

$$q_{n,d,\sigma} = a + \min\{u, s + t\}.$$

For nonnegative a, t, u, s with $t \leq u$, the lower bound follows by splitting according as $u \leq s + t$ or $s + t < u$, while the upper bound follows by splitting according as $u \leq s$ or $s < u$. These two checks give

$$\frac{1}{2}\{a + t + \min(u, s)\} \leq a + \min\{u, s + t\} \leq a + t + \min(u, s).$$

Thus

$$\frac{1}{2}\left\{b_{n,d} + \min(u_{n,d}, \sigma^2)\right\} \leq q_{n,d,\sigma} \leq b_{n,d} + \min(u_{n,d}, \sigma^2),$$

and the product benchmark is, by definition,

$$\ell_{n,d,\sigma} = b_{n,d} + \sigma^2 u_{n,d}.$$

For every radius floor $\underline{\sigma} > 0$, put

$$D_{\underline{\sigma}} := \max\{1, \underline{\sigma}^{-2}\}.$$

If $\underline{\sigma} \leq \sigma \leq 2$, then $\sigma^2 \geq \underline{\sigma}^2$, and the frontier and converse benchmarks satisfy

$$r_{n,d,\sigma} \leq \frac{1}{n} + \min(1, u_{n,d}) \leq D_{\underline{\sigma}}\left\{\frac{1}{n} + \sigma^2 \min(1, u_{n,d})\right\} \leq D_{\underline{\sigma}}\omega_{n,d,\sigma}.$$

Combining this comparison with Theorem 4, which gives

$$c_0 M^2 \omega_{n,d,\sigma} \leq \mathsf{R}_{n,d,\epsilon,M,\sigma} \leq C_0 M^2 r_{n,d,\sigma},$$

yields the fixed-radius constants $c = c_0/D_{\underline{\sigma}}$ and $C = C_0$:

$$cM^2 r_{n,d,\sigma} \leq \mathsf{R}_{n,d,\epsilon,M,\sigma} \leq CM^2 r_{n,d,\sigma}.$$

For the matching elbows, the benchmark comparison used with the same bracket is

$$r_{n,d,\sigma} \leq (1 + K)\omega_{n,d,\sigma}.$$

When $u_{n,d} \geq 1$,

$$r_{n,d,\sigma} \leq \frac{1}{n} + \min\left\{1, \sigma^2 + \frac{d}{n^2}\right\} \leq \omega_{n,d,\sigma},$$

where the last inequality is checked by the two cases $d/n^2 \leq 1$ and $d/n^2 > 1$. When $u_{n,d} \leq Kb_{n,d}$, the case $d/n^2 \leq 1$ gives

$$r_{n,d,\sigma} \leq \frac{1}{n} + u_{n,d} \leq (1 + K)b_{n,d} \leq (1 + K)\omega_{n,d,\sigma},$$

and the case $d/n^2 > 1$ gives $r_{n,d,\sigma} \le n^{-1} + 1 \le \omega_{n,d,\sigma}$. When $\sigma^2 \le K b_{n,d}$, the case $d/n^2 \le 1$ gives

$$r_{n,d,\sigma} \le b_{n,d} + \sigma^2 \le (1+K)b_{n,d} \le (1+K)\omega_{n,d,\sigma},$$

and the case $d/n^2 > 1$ is again $r_{n,d,\sigma} \le n^{-1}+1 \le \omega_{n,d,\sigma}$. The all-alphabet bracket therefore gives the matching-elbow minimax equivalence with lower constant $c_0/(1+K)$ and upper constant C_0.

For $n > 0$, the logarithmic factor is positive. Expanding $u_{n,d} \le b_{n,d}$, multiplying by the positive denominator $n^2L_n^2$, and collecting terms gives

$$u_{n,d} \le b_{n,d} \quad \Longleftrightarrow \quad d^2 \le nL_n^2 + dL_n^2.$$

The sufficient elbows are the two displayed consequences of this algebra. If $d \le \sqrt{n}\,L_n$, then $d^2 \le nL_n^2$; if $d \le L_n^2$, then $d^2 \le dL_n^2$. In either case the exact algebraic elbow gives $u_{n,d} \le b_{n,d}$.

For the residual-wedge implication, write along the sequence $b_j = b_{j,d_j}$, $u_j = u_{j,d_j}$, $s_j = \sigma_j^2$, $r_j = r_{j,d_j,\sigma_j}$, and $\omega_j = \omega_{j,d_j,\sigma_j}$. Since $r_j/\omega_j \to \infty$, the matching-elbow comparison implies that, for every fixed $K \ge 0$, eventually

$$u_j < 1, \qquad Kb_j < u_j, \qquad Kb_j < s_j.$$

Taking $K = \eta^{-1}$ for arbitrary $\eta > 0$ gives $b_j/u_j \to 0$ and $b_j/s_j \to 0$. To prove $s_j \to 0$, fix $\alpha > 0$, put $\rho = \sqrt{\alpha}$, and set $D = \max\{1, \rho^{-2}\}$. Divergence makes $r_j/\omega_j > D$ eventually; on the same tail, $\sigma_j \ge \rho$ would give $r_j \le D\omega_j$ by the fixed-radius comparison. Hence $0 \le \sigma_j < \rho$ eventually, so $s_j < \alpha$ eventually.

It remains to show $u_j \to 0$. Fix $\alpha > 0$. If $\alpha \ge 1$, the eventual inequality $u_j < 1$ gives $u_j < \alpha$. If $0 < \alpha < 1$, use the eventual bounds $b_j/u_j < 1/2$, $s_j < \alpha/2$, $u_j < 1$, and $r_j/\omega_j > \alpha^{-1}$. At any point on this tail with $u_j \ge \alpha$, the inequality $d_j/j^2 \le b_j < u_j/2$ and $s_j < u_j/2$ give $s_j + d_j/j^2 < u_j$ and $d_j/j^2 \le 1$. Consequently

$$r_j = b_j + s_j, \qquad \omega_j = b_j + s_ju_j,$$

and $u_j \ge \alpha$ with $\alpha < 1$ implies $r_j/\omega_j \le \alpha^{-1}$, a contradiction. Thus $u_j < \alpha$ eventually. Together with the assumed eventual admissibility, these four limits are exactly the residual-wedge conditions.

For the diagonal sequence $d_n = n$ and $\sigma_n = L_n^{-1/2}$, eventual admissibility follows from $L_n \ge 1$. Direct simplification gives, for $n \ge 1$,

$$b_{n,n} = \frac{2}{n}, \qquad u_{n,n} = \frac{1}{L_n^2}, \qquad \sigma_n^2 = \frac{1}{L_n}.$$

The logarithmic facts $L_n \to \infty$ and $L_n^k/n \to 0$ for each fixed k yield

$$\frac{b_{n,n}}{u_{n,n}} = \frac{2L_n^2}{n} \to 0, \qquad u_{n,n} = \frac{1}{L_n^2} \to 0, \qquad \frac{b_{n,n}}{\sigma_n^2} = \frac{2L_n}{n} \to 0, \qquad \sigma_n^2 = \frac{1}{L_n} \to 0.$$

Hence the diagonal sequence lies in the residual wedge.

Along the same sequence, the exact benchmark formulas are

$$q_{n,n,L_n^{-1/2}} = \frac{1}{n} + \frac{1}{L_n^2}, \qquad \ell_{n,n,L_n^{-1/2}} = \frac{2}{n} + \frac{1}{L_n^3}.$$

Since $L_n^3/n \to 0$, these formulas give eventually

$$\frac{1}{3}L_n \le \frac{q_{n,n,L_n^{-1/2}}}{\ell_{n,n,L_n^{-1/2}}} \le 2L_n.$$

Thus the diagonal benchmark ratio is of order L_n, and because $L_n \to \infty$, the ratio diverges. $\square$

Proof of Theorem 4. 1. Fix an overlap parameter ϵ with $0 < \epsilon < 1/2$. By Theorem 2, choose a constant $C_\epsilon^{\mathrm{up}} > 0$ and a polynomial calibration handle such that, for every $n, d \ge 1$, $M \ge 1$, and $0 \le \sigma \le 2$, the corresponding known-radius selector $\widehat{\tau}_n^\star \in \mathcal{T}_{n,M}$ satisfies

$$\sup_{P \in \mathcal{P}_{d,\epsilon,M,\sigma}} E_{Q_P^{(n)}}\left[\{\widehat{\tau}_n^\star - \tau(P)\}^2\right] \le C_\epsilon^{\mathrm{up}} M^2 r_{n,d,\sigma}.$$

By Theorem 3, choose c_ϵ^{lo} and $b_\epsilon^{\mathrm{rad}}$ such that

$$0 < c_\epsilon^{\mathrm{lo}} \le 1, \qquad 0 < b_\epsilon^{\mathrm{rad}},$$

and, for every such n, d, M, σ,

$$c_\epsilon^{\mathrm{lo}} M^2 \ell_{n,d,\sigma} \le \mathsf{R}_{n,d,\epsilon,M,\sigma}.$$

Set

$$c_\epsilon := c_\epsilon^{\mathrm{lo}}, \qquad C_\epsilon := \max\{1, C_\epsilon^{\mathrm{up}}\}.$$

Then

$$0 < c_\epsilon \le 1, \qquad 1 \le C_\epsilon, \qquad c_\epsilon \le C_\epsilon.$$

2. Fix $n, d \ge 1$, $M \ge 1$, and $0 \le \sigma \le 2$. The lower bound is exactly the one supplied by Theorem 3 with the constants fixed above:

$$c_\epsilon M^2 \ell_{n,d,\sigma} = c_\epsilon^{\mathrm{lo}} M^2 \ell_{n,d,\sigma} \le \mathsf{R}_{n,d,\epsilon,M,\sigma}.$$

The accompanying radius-channel certificate in Theorem 3 also supplies a realized source family and an embedding into the same model class. Applying its source-risk assertion to the zero source estimator gives a source law, and the embedding-membership part of the certificate gives an embedded law $P_\bullet$ with

$$P_\bullet \in \mathcal{P}_{d,\epsilon,M,\sigma}.$$

Thus the model class is nonempty for these indices.

3. For any estimator $\widehat{\eta} \in \mathcal{T}_{n,M}$, define its worst-case risk over this class by

$$\mathcal{W}_{n,d,\epsilon,M,\sigma}(\widehat{\eta}) := \sup_{P \in \mathcal{P}_{d,\epsilon,M,\sigma}} E_{Q_P^{(n)}}\left[\{\widehat{\eta} - \tau(P)\}^2\right].$$

We first record the lower-boundedness needed to apply the infimum in Definition 5. Fix $\widehat{\eta} \in \mathcal{T}_{n,M}$. In the vacuous branch where the model index is empty, the real supremum convention gives $\mathcal{W}_{n,d,\epsilon,M,\sigma}(\widehat{\eta}) = 0$. In the nonvacuous branch, consider the range of mean-squared errors indexed by $P \in \mathcal{P}_{d,\epsilon,M,\sigma}$. If this range is bounded above, choose any law in the class; its mean-squared error is the integral of a square and is therefore nonnegative, and it is bounded by the supremum defining $\mathcal{W}_{n,d,\epsilon,M,\sigma}(\widehat{\eta})$. If the range is unbounded above, the supremum is in its unbounded-above case and the inequality $0 \le \mathcal{W}_{n,d,\epsilon,M,\sigma}(\widehat{\eta})$ is immediate. Hence

$$0 \le \mathcal{W}_{n,d,\epsilon,M,\sigma}(\widehat{\eta}) \qquad \text{for every } \widehat{\eta} \in \mathcal{T}_{n,M}.$$

The collection of worst-case risks is therefore bounded below by 0. Using Definition 5 and the admissibility of $\widehat{\tau}_n^\star$,

$$\mathsf{R}_{n,d,\epsilon,M,\sigma} = \inf_{\widehat{\eta} \in \mathcal{T}_{n,M}} \mathcal{W}_{n,d,\epsilon,M,\sigma}(\widehat{\eta}) \le \mathcal{W}_{n,d,\epsilon,M,\sigma}(\widehat{\tau}_n^\star).$$

4. Applying the selector guarantee from Theorem 2 gives

$$\mathcal{W}_{n,d,\epsilon,M,\sigma}(\widehat{\tau}_n^\star) \le C_\epsilon^{\mathrm{up}} M^2 r_{n,d,\sigma}.$$

Moreover,

$$u_{n,d} \ge 0, \qquad h_{n,d,\sigma} \ge 0,$$

so

$$r_{n,d,\sigma} = \frac{1}{n} + \min\{1, u_{n,d}, h_{n,d,\sigma}\} \ge \frac{1}{n} > 0.$$

Therefore $M^2 r_{n,d,\sigma} \ge 0$, and $C_\epsilon^{\mathrm{up}} \le C_\epsilon$ implies

$$C_\epsilon^{\mathrm{up}} M^2 r_{n,d,\sigma} \le C_\epsilon M^2 r_{n,d,\sigma}.$$

Combining the previous displays yields

$$\mathsf{R}_{n,d,\epsilon,M,\sigma} \le C_\epsilon M^2 r_{n,d,\sigma}.$$

Together with the lower bound supplied by Theorem 3 and instantiated above, this proves the asserted bracket for all $n, d \ge 1$, $M \ge 1$, and $0 \le \sigma \le 2$.

□

Proof of Proposition 3. Fix $\epsilon \in (0, 1/2)$ and assume the published binary collision guarantee stated in the proposition. We prove the simultaneous conclusion in three steps.

1. The assumed guarantee is exactly the first component of the simultaneous assertion: it supplies the stated constants $C_\epsilon, b_\epsilon > 0$ and the two displayed bias and variance bounds for every admissible n, d, P, and σ_{bin}.

2. For the pointwise algebraic comparison, fix $n, d \in \mathbb{N}$ and $\sigma \in \mathbb{R}$ in the stated range. The defining order property of the minimum gives

$$\min\{1, \min\{u_{n,d}, h_{n,d,\sigma}\}\} \le \min\{u_{n,d}, h_{n,d,\sigma}\} \le h_{n,d,\sigma}.$$

This is the asserted pointwise inequality.

3. For the asymptotic comparison, let d_n and σ_n satisfy the stated sequence hypotheses. Limits along $\mathbb{N}$ are unchanged by discarding finitely many initial indices, so work on the cofinal set $n > 0$. There $d_n > 0$, and hence

$$h_{n,d_n,\sigma_n} = \sigma_n^2 + \frac{d_n}{n^2} > 0.$$

Also $u_{n,d_n} \geq 0$, $h_{n,d_n,\sigma_n} \geq 0$, and $1 \geq 0$. Therefore

$$0 \leq \frac{\min\{1, \min\{u_{n,d_n}, h_{n,d_n,\sigma_n}\}\}}{h_{n,d_n,\sigma_n}}.$$

The same minimum inequalities, now using the left branch of the inner minimum, give

$$\min\{1, \min\{u_{n,d_n}, h_{n,d_n,\sigma_n}\}\} \leq \min\{u_{n,d_n}, h_{n,d_n,\sigma_n}\} \leq u_{n,d_n}.$$

Dividing by the positive denominator h_{n,d_n,σ_n} yields, for all sufficiently large n,

$$0 \leq \frac{\min\{1, \min\{u_{n,d_n}, h_{n,d_n,\sigma_n}\}\}}{h_{n,d_n,\sigma_n}} \leq \frac{u_{n,d_n}}{h_{n,d_n,\sigma_n}}.$$

By hypothesis the right-hand side converges to 0, so the squeeze theorem gives

$$\frac{\min\{1, \min\{u_{n,d_n}, h_{n,d_n,\sigma_n}\}\}}{h_{n,d_n,\sigma_n}} \to 0.$$

This proves the sequence comparison and completes the proof.

□

Proof of Proposition 1. Fix $0 < \epsilon < 1/2$. Write $\bar{c}_{\text{ex}}$ for the constant supplied by Lemma 2, take c_{rad} and b_{rad} from Theorem 3, and take C_{up} and the polynomial handle from Lemma 13. Let

$$u_\epsilon := \min\{1/2, 8(1/2 - \epsilon)^2\}$$

and set

$$c_{\text{ex}} := \min\left\{\bar{c}_{\text{ex}}, \frac{1}{4}\min\left\{\frac{1}{100}, \frac{u_\epsilon^2}{64}\right\}, 1\right\}.$$

Then $c_{\text{ex}} > 0$, $c_{\text{ex}} \leq \bar{c}_{\text{ex}}$, $c_{\text{ex}} \leq 1$, and

$$c_{\text{ex}} \leq \frac{1}{4}\min\left\{\frac{1}{100}, \frac{u_\epsilon^2}{64}\right\}.$$

The selected polynomial handle retains its cutoff and active-range multiplier from Lemma 13.

For each $d \geq 1$ and $M \geq 1$, define Φ_M to be the affine binary-to-real pushforward of Definition 8, namely $B(a) \mapsto M\{B(a) - 1/2\}$. The embedding identities preserve cell masses and propensities, push each positive arm-cell binary outcome law through this affine map, and make each arm-cell mean equal to M times the corresponding binary mean minus $M/2$. Hence every binary law satisfying the overlap condition gives a member of $\mathcal{P}^{\text{unr}}_{d,\epsilon,M}$ as in Definition 2, and every uniform-mass exactly homogeneous binary law gives a member of the zero-radius class,

$$\Phi_M(P) \in \mathcal{P}_{d,\epsilon,M,0}.$$

Fix a cell k in the nonempty alphabet and fix $0 \leq \sigma \leq 2$. Let Q be the one-cell test law that puts unit cell mass on k, has propensity $1/2$, has control arm-cell outcome law δ_0, and has treated arm-cell outcome law given by the two-point mean-zero law on $\{-M/2, M/2\}$. With local mean parameter zero, this test law belongs to $\mathcal{P}_{d,\epsilon,M,\sigma}$. For any binary law P, the control arm-cell outcome law of $\Phi_M(P)$ is supported on the two affine binary values $\{-M/2, M/2\}$, which assign zero mass to $\{0\}$ since $M \geq 1$. The corresponding control arm-cell law of Q assigns mass one to $\{0\}$. Therefore

$$\Phi_M(P) \neq Q$$

for every binary law P, giving the strict-image assertion at radius σ.

Now fix $n \geq 1$ and $0 \leq \sigma \leq 2$, and assume $d \leq c_{\mathrm{ex}} n^2$. Since $c_{\mathrm{ex}} \leq 1$, this gives $d \leq n^2$. For the binary source experiment in Definition 10, a two-point subexperiment supplies the parametric term, as follows. Put $\delta := (2/5)/\sqrt{n}$, so that $\delta \geq 0$, $\delta^2 = 4/(25n)$, and $1/2 + \delta \leq 1$. Let P_0, P_1 be the two uniform-mass exactly homogeneous binary laws with constant propensity $1/2$, control-arm response mean $1/2$ under both, and treated-arm response mean $1/2$ under P_0 and $1/2 + \delta$ under P_1; constant propensity $1/2$ meets the overlap constraint because $0 < \epsilon < 1/2$, and both laws are exactly homogeneous because the cell contrast is the same in every cell. Their targets differ by $|\psi(P_0) - \psi(P_1)| = \delta$, and the product-law calibration

$$n \, \frac{(1/2)\delta^2}{(1/2)(1 - 1/2)} = \frac{8}{25} \leq \log 2$$

gives $\mathrm{TV}(P_0^{\otimes n}, P_1^{\otimes n}) \leq 1/2$. The two-point testing bound, with centres $1/2$ and $1/2+\delta$, threshold $\delta/2$, and total-variation constant $1/2$, therefore yields

$$\frac{1}{100n} \leq \inf_{\widehat{\psi}} \sup_{P \in \mathcal{H}^{\mathrm{bin}}_{n,d}(\epsilon)} E_{P^{\otimes n}} \left[\{\widehat{\psi} - \psi(P)\}^2 \right],$$

and Lemma 3 gives

$$\frac{u_\epsilon^2}{64} \frac{d}{n^2} \leq \inf_{\widehat{\psi}} \sup_{P \in \mathcal{H}^{\mathrm{bin}}_{n,d}(\epsilon)} E_{P^{\otimes n}} \left[\{\widehat{\psi} - \psi(P)\}^2 \right].$$

Together with the definition of c_{ex}, these two inequalities place

$$c_{\mathrm{ex}} \left(\frac{1}{n} + \frac{d}{n^2} \right)$$

strictly below the fixed-sample binary exact-homogeneity minimax risk. Thus every measurable binary estimator has a source law $P \in \mathcal{H}^{\mathrm{bin}}_{n,d}(\epsilon)$ whose binary squared risk is at least this level. Applying this statement to the binary estimator obtained by composing an arbitrary clipped real-outcome estimator with the affine observation map, and using $\tau(\Phi_M(P)) = M\psi(P)$ on the exact-homogeneous source class, multiplies the lower bound by M^2. Hence every such real-outcome estimator has some $P \in \mathcal{H}^{\mathrm{bin}}_{n,d}(\epsilon)$ satisfying

$$c_{\mathrm{ex}} M^2 \left(\frac{1}{n} + \frac{d}{n^2} \right) \leq \mathrm{MSE}_{\Phi_M(P)}(\widehat{\tau}).$$

The exact-range minimax-risk bound follows from Lemma 2. Since $c_{\mathrm{ex}} \leq \bar{c}_{\mathrm{ex}}$ and $d \leq n^2$, the term $\min\{1, d/n^2\}$ in that all-alphabet transfer equals d/n^2 on the present range, giving

$$c_{\mathrm{ex}} M^2 \left(\frac{1}{n} + \frac{d}{n^2} \right) \leq \mathsf{R}_{n,d,\epsilon,M,\sigma}.$$

For the radius term, Theorem 3 supplies

$$c_{\mathrm{rad}} M^2 \left\{ \frac{1}{n} + \min\left(1, \frac{d}{n^2}\right) + \sigma^2 \min\{1, u_{n,d}\} \right\} \le \mathsf{R}_{n,d,\epsilon,M,\sigma},$$

together with the transfer certificate with constants c_{rad} and b_{rad}. The base summand $1/n + \min\{1, d/n^2\}$ is nonnegative, so the same inequality yields

$$c_{\mathrm{rad}} M^2 \sigma^2 \min\{1, u_{n,d}\} \le \mathsf{R}_{n,d,\epsilon,M,\sigma}.$$

Finally, Lemma 13 supplies the selected polynomial handle. The associated estimator T_n^{poly} is measurable, takes values in $[-M, M]$, satisfies the handle's active-range implication for K_n, and obeys uniformly for every $Q \in \mathcal{P}_{d,\epsilon,M,\sigma}$

$$\mathrm{MSE}_Q(T_n^{\mathrm{poly}}) \le C_{\mathrm{up}} M^2 \left(\frac{1}{n} + \min\{1, u_{n,d}\} \right).$$

□

Proof of Theorem 3. Fix $0 < \epsilon < 1/2$. Apply Lemma 5 to obtain one-arm constants $a_{\mathrm{rad}}, b_0, N_0$ with $a_{\mathrm{rad}} > 0$ and $b_0 > 0$. Put

$$\tilde{b}_0 := \min\{b_0, 4\}, \qquad b_\epsilon^{\mathrm{rad}} := \tilde{b}_0/2, \qquad \bar{b}_\epsilon^{\mathrm{rad}} := b_0, \qquad c_{\mathrm{rad}} := a_{\mathrm{rad}} \tilde{b}_0^2/128,$$

and choose

$$N_{\mathrm{rad},0} := \max\{3, N_0, \lceil 2/\tilde{b}_0 \rceil\}.$$

Then

$$0 < a_{\mathrm{rad}}, \qquad 0 < b_\epsilon^{\mathrm{rad}} < \bar{b}_\epsilon^{\mathrm{rad}}, \qquad 0 < c_{\mathrm{rad}}.$$

For every $n, d \ge 1$ with $N_{\mathrm{rad},0} \le n$, the radial cap

$$d_{\mathrm{rad}} = \min\{d, \max(1, \lfloor b_\epsilon^{\mathrm{rad}} n \log(en) \rfloor)\}$$

is positive, is at most d, and satisfies

$$d_{\mathrm{rad}} \le \bar{b}_\epsilon^{\mathrm{rad}} n \log n.$$

The same construction gives the source-risk interface used by the channel transport: for every measurable uniformly bounded source estimator $\widehat{\psi}$ on the n-sample binary one-arm experiment over d_{rad} cells, some control-zero source law P satisfies

$$\frac{a_{\mathrm{rad}}}{2} \left\{ \frac{1}{n} + \frac{d_{\mathrm{rad}}^2}{n^2 \log^2 n} \right\} \le E_{Q_P^{(n)}} \left[\{\widehat{\psi} - \psi(P)\}^2 \right].$$

Finally, Lemma 16, applied with cap constant $\tilde{b}_0 = 2 b_\epsilon^{\mathrm{rad}}$, gives for every M and σ

$$c_{\mathrm{rad}} M^2 \sigma^2 \min\left(1, \frac{d^2}{n^2 \log^2(en)}\right) \le \left(\frac{M\sigma}{2}\right)^2 \frac{a_{\mathrm{rad}}}{2} \left\{ \frac{1}{n} + \frac{d_{\mathrm{rad}}^2}{n^2 \log^2 n} \right\}.$$

Choose $b_\epsilon^{\mathrm{ex}}, c_{\mathrm{ex}} > 0$ and N_{ex} from Lemma 15. Let $c_{\mathrm{base}} > 0$ be the constant supplied by Lemma 2.

Set

$$N_{\text{rad}} := \max\{N_{\text{rad},0}, 1\}, \qquad c_{\text{small}} := \frac{1}{800 N_{\text{rad}}}, \qquad c_{\text{tr}} := \min\{c_{\text{rad}}, c_{\text{small}}, c_{\text{ex}}\},$$

and define

$$c_\epsilon := \min\left\{1, \frac{1}{2}\min\{c_{\text{base}}, c_{\text{tr}}\}\right\}.$$

All constants in this display depend on ϵ alone, and the displayed positivity gives $0 < c_\epsilon \le 1$ and $b_\epsilon^{\text{rad}} > 0$.

Fix n, d, M, σ with $0 < n$, $0 < d$, $1 \le M$, and $0 \le \sigma \le 2$. Form the two capped alphabets d_{rad} and

$$d_{\text{ex}} = \min\{d, \max(1, \lfloor b_\epsilon^{\text{ex}} n^2 \rfloor)\}.$$

Both caps are positive and at most d. Following Algorithm 4, pad each source alphabet into the first target cells. On the radial source, apply the hypothesis-independent Bernoulli channel

$$B'(a) \mid B(a) \sim \text{Bernoulli}\left\{\frac{1}{2} + \frac{\sigma}{2}\left(B(a) - \frac{1}{2}\right)\right\}, \qquad Y(a) = M\{B'(a) - 1/2\}.$$

On the exact-homogeneity source, use the affine map $Y(a) = M\{B(a) - 1/2\}$. The construction preserves the source cell masses on the padded cells, assigns zero mass outside them, realizes the independent full-data couplings, and gives the radial full-data channel of Definition 12. For a radial source law P, write $\eta_k = E_P[B(1) \mid A = 1, X = k]$ on a positive source cell and $\psi(P) = \sum_k p_k \eta_k$. On each padded positive-mass cell the embedded means are

$$\mu_{0k} = -\frac{M\sigma}{4}, \qquad \mu_{1k} = \frac{M\sigma}{2}\left(\eta_k - \frac{1}{2}\right), \qquad \tau_k = \mu_{1k} - \mu_{0k} = \frac{M\sigma}{2}\eta_k,$$

and

$$\tau(\Psi(P)) = \frac{M\sigma}{2}\psi(P).$$

Thus the overlap bounds are inherited from the source, the mean-normalization bounds in Assumption 4 follow from $0 \le \eta_k \le 1$ and $0 \le \sigma \le 2$, the conditional second-moment bound in Assumption 6 follows from $Y(a) \in \{-M/2, M/2\}$, and the radius condition in Assumption 5 follows from

$$|\tau_k - \tau(\Psi(P))| = \frac{M\sigma}{2}|\eta_k - \psi(P)| \le \sigma M.$$

Together with consistency and conditional exchangeability from the independent full-data coupling, these checks place every radial embedding in $\mathcal{P}_{d,\epsilon,M,\sigma}$ as defined in Definition 1. Along the same embedding,

$$\tau(\Psi(P_1)) - \tau(\Psi(P_0)) = M\frac{\sigma}{2}\{\psi(P_1) - \psi(P_0)\}.$$

The exact affine embedding has target slope M.

The exact capped package supplies, for this (n, d), a source level L_{ex} such that every measurable exact-source estimator is defeated by some exactly homogeneous binary source law and

$$c_{\text{ex}} M^2 \left\{\frac{1}{n} + \min\left(1, \frac{d}{n^2}\right)\right\} \le M^2 L_{\text{ex}}.$$

Since $c_{\mathrm{tr}} \le c_{\mathrm{ex}}$, deterministic affine transport gives, for every $\widehat{\tau} \in \mathcal{T}_{n,M}$, an exact source law P satisfying

$$c_{\mathrm{tr}} M^2 \left\{ \frac{1}{n} + \min\left(1, \frac{d}{n^2}\right) \right\} \le \mathrm{MSE}_{\Phi_M(P)}(\widehat{\tau}),$$

where MSE is the mean-squared error of Definition 9. This is the exact half of the common two-family transfer certificate.

For the radial half, let K be the one-observation Markov kernel that maps a binary observed record (X, A, B) to $(\mathrm{pad}(X), A, M(B' - 1/2))$ using the Bernoulli channel above. For every radial source law P, the observed target law satisfies

$$(\Psi(P))_O = K \circ P_O,$$

and the target identity is

$$\tau(\Psi(P)) = \frac{M\sigma}{2} \psi(P).$$

Consequently, whenever $\sigma > 0$, the nonzero scale $M\sigma/2$, the preceding observed-law identity, and the source-risk interface imply the following transport rule: if for every measurable uniformly bounded $\widehat{\psi}$ some source law P satisfies

$$L \le E_{Q_P^{(n)}} \left[\{\widehat{\psi} - \psi(P)\}^2 \right]$$

and if

$$c\, M^2 \sigma^2 \min\left(1, \frac{d^2}{n^2 \log^2(en)}\right) \le \left(\frac{M\sigma}{2}\right)^2 L,$$

then every $\widehat{\tau} \in \mathcal{T}_{n,M}$ has some radial source law P with

$$c\, M^2 \sigma^2 \min\left(1, \frac{d^2}{n^2 \log^2(en)}\right) \le \mathrm{MSE}_{\Psi(P)}(\widehat{\tau}).$$

This is the Markov-kernel comparison applied to the radial channel. The finite-sample source interface used when the radial cutoff is unavailable is obtained from the one-arm parametric subexperiment. In the binary one-arm experiment over d_{rad} cells, take the two control-zero source laws whose treated functionals are $1/2$ and $1/2 + \delta$, where $\delta = (2/5)/\sqrt{n}$. The usual product total-variation calculation gives distance at most $1/2$ between their n-sample observed laws, so the standard two-point minimax inequality gives

$$\frac{1}{100n} \le \inf_{\widetilde{\psi}} \sup_{R} E_{Q_R^{(n)}} \left[\{\widetilde{\psi} - \psi(R)\}^2 \right],$$

where R ranges over the control-zero one-arm source laws on d_{rad} cells. Since $1/(200n)$ is strictly below this minimax value, the hard-source extraction from the minimax infimum and supremum gives, for every measurable uniformly bounded source estimator $\widehat{\psi}$, a control-zero source law P with

$$\frac{1}{200n} \le E_{Q_P^{(n)}} \left[\{\widehat{\psi} - \psi(P)\}^2 \right].$$

If $\sigma = 0$, the desired radial lower-bound term is zero; this finite-sample source law and nonnegativity of squared loss give the required estimator-wise inequality. If $\sigma > 0$ and $N_{\mathrm{rad},0} \le n$, the large-sample source-risk interface and the scale comparison displayed above give

$$c_{\mathrm{tr}} M^2 \sigma^2 \min\left(1, \frac{d^2}{n^2 \log^2(en)}\right) \le \left(\frac{M\sigma}{2}\right)^2 \frac{a_{\mathrm{rad}}}{2} \left\{ \frac{1}{n} + \frac{d_{\mathrm{rad}}^2}{n^2 \log^2 n} \right\}.$$

The transport rule therefore gives the radial estimator-wise lower bound in the large-sample branch. If $\sigma > 0$ and $n < N_{\mathrm{rad},0}$, then $1 \le n \le N_{\mathrm{rad}}$. The finite-sample source interface just displayed, together with Lemma 17 and $c_{\mathrm{tr}} \le c_{\mathrm{small}}$, gives the transported radial lower term in the small-sample branch. Hence for every $\widehat{\tau} \in \mathcal{T}_{n,M}$ there is a radial source law P with

$$c_{\mathrm{tr}} M^2 \sigma^2 \min\left(1, \frac{d^2}{n^2 \log^2(en)}\right) \le \mathrm{MSE}_{\Psi(P)}(\widehat{\tau}).$$

Combining the radial and exact estimator-wise statements gives the common transfer certificate with constant c_{tr}. Since $c_\epsilon \le c_{\mathrm{tr}}$, the same certificate holds with c_ϵ. The radial construction also carries the product data-processing certificate: writing TV for total variation distance, for all radial source laws P_0, P_1,

$$\mathrm{TV}\left(Q^{(n)}_{\Psi(P_0)}, Q^{(n)}_{\Psi(P_1)}\right) \le \mathrm{TV}\left(Q^{(n)}_{P_0}, Q^{(n)}_{P_1}\right),$$

because the left product laws are obtained from the right product laws by the coordinatewise common kernel K. Thus the cap identity, injective padding, source hardness, mass preservation, zero extension, full-data coupling, full-data channel law, membership in $\mathcal{P}_{d,\epsilon,M,\sigma}$, radius bound, product data-processing bound, target scaling, and radial estimator-wise lower bound package into the radial certificate asserted in the theorem.

For the minimax lower bound, Lemma 2 gives

$$c_{\mathrm{base}} M^2 \left\{ \frac{1}{n} + \min\left(1, \frac{d}{n^2}\right) \right\} \le \mathsf{R}_{n,d,\epsilon,M,\sigma}.$$

The radial half of the transfer certificate, after embedding into $\mathcal{P}_{d,\epsilon,M,\sigma}$ and taking the minimax infimum in Definition 5, gives

$$c_{\mathrm{tr}} M^2 \sigma^2 \min\left(1, \frac{d^2}{n^2 \log^2(en)}\right) \le \mathsf{R}_{n,d,\epsilon,M,\sigma}.$$

For two nonnegative terms bounded above by the same risk in this way,

$$\frac{1}{2} \min\{c_{\mathrm{base}}, c_{\mathrm{tr}}\} M^2 \left\{ \frac{1}{n} + \min\left(1, \frac{d}{n^2}\right) + \sigma^2 \min\left(1, \frac{d^2}{n^2 \log^2(en)}\right) \right\} \le \mathsf{R}_{n,d,\epsilon,M,\sigma}.$$

The definition of c_ϵ yields the displayed converse rate with constant c_ϵ, completing the proof. $\square$

References

Alexandre Belloni, Victor Chernozhukov, Iván Fernández-Val, and Christian Hansen. Program evaluation and causal inference with high-dimensional data. *Econometrica*, 85(1):233–298, 2017. doi: 10.3982/ECTA12723.

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.

Michael Celentano and Martin J. Wainwright. Challenges of the inconsistency regime: Novel debiasing methods for missing data models, 2023.

Victor Chernozhukov, Denis Chetverikov, Mert Demirer, Esther Duflo, Christian Hansen, Whitney Newey, and James Robins. Double/debiased machine learning for treatment and structural parameters. *The Econometrics Journal*, 21(1):C1–C68, 2018. doi: 10.1111/ectj.12097.

David L. Donoho, Richard C. Liu, and Brenda MacGibbon. Minimax risk over hyperrectangles, and implications. *The Annals of Statistics*, 18(3):1416–1437, 1990. doi: 10.1214/aos/1176347758.

Jinyong Hahn. On the role of the propensity score in efficient semiparametric estimation of average treatment effects. *Econometrica*, 66(2):315–331, 1998. doi: 10.2307/2998560.

Keisuke Hirano, Guido W. Imbens, and Geert Ridder. Efficient estimation of average treatment effects using the estimated propensity score. *Econometrica*, 71(4):1161–1189, 2003. doi: 10.1111/1468-0262.00442.

Wassily Hoeffding. A class of statistics with asymptotically normal distribution. *The Annals of Mathematical Statistics*, 19(3):293–325, 1948. doi: 10.1214/aoms/1177730196.

Guido W. Imbens and Donald B. Rubin. *Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction*. Cambridge University Press, Cambridge, 2015. ISBN 9780521885881.

Guido W. Imbens and Jeffrey M. Wooldridge. Recent developments in the econometrics of program evaluation. *Journal of Economic Literature*, 47(1):5–86, 2009. doi: 10.1257/jel.47.1.5.

Kuanhao Jiang, Rajarshi Mukherjee, Subhabrata Sen, and Pragya Sur. A new central limit theorem for the augmented ipw estimator: Variance inflation, cross-fit covariance and beyond. *The Annals of Statistics*, 53(2), 2025. doi: 10.1214/24-AOS2476.

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.

Edward H. Kennedy. Semiparametric doubly robust targeted double machine learning: A review, 2022.

Edward H. Kennedy. Towards optimal doubly robust estimation of heterogeneous causal effects. *Electronic Journal of Statistics*, 17(2):3008–3049, 2023. doi: 10.1214/23-EJS2157.

Lucien Le Cam. *Asymptotic Methods in Statistical Decision Theory*. Springer, New York, 1986. ISBN 9780387963075. doi: 10.1007/978-1-4612-4946-7.

Gábor Lugosi and Shahar Mendelson. Sub-gaussian estimators of the mean of a random vector. *The Annals of Statistics*, 47(2):783–794, 2019. doi: 10.1214/17-AOS1639.

Xinkun Nie and Stefan Wager. Quasi-oracle estimation of heterogeneous treatment effects. *Biometrika*, 108(2):299–319, 2021. doi: 10.1093/biomet/asaa076.

Liam Paninski. Estimation of entropy and mutual information. *Neural Computation*, 15(6):1191–1253, 2003. doi: 10.1162/089976603321780272.

James M. Poterba, Steven F. Venti, and David A. Wise. 401(k) plans and tax-deferred saving. In David A. Wise, editor, *Studies in the Economics of Aging*, pages 105–142. University of Chicago Press, Chicago, 1994. URL https://www.nber.org/books-and-chapters/studies-economics-aging/401k-plans-and-tax-deferred-saving.

James M. Poterba, Steven F. Venti, and David A. Wise. Do 401(k) contributions crowd out other personal saving? *Journal of Public Economics*, 58(1):1–32, 1995. doi: 10.1016/0047-2727(94)01462-W.

James M. Robins, Andrea Rotnitzky, and Lue Ping Zhao. Estimation of regression coefficients when some regressors are not always observed. *Journal of the American Statistical Association*, 89(427):846–866, 1994. doi: 10.1080/01621459.1994.10476818.

James M. Robins, Lingling Li, Eric Tchetgen, and Aad W. van der Vaart. Higher order influence functions and minimax estimation of nonlinear functionals. In *Probability and Statistics: Essays in Honor of David A. Freedman*, volume 2 of *Institute of Mathematical Statistics Collections*, pages 335–421. Institute of Mathematical Statistics, Beachwood, OH, 2008.

James M. Robins, Lingling Li, Rajarshi Mukherjee, Eric Tchetgen Tchetgen, and Aad W. van der Vaart. Minimax estimation of a functional on a structured high-dimensional model. *The Annals of Statistics*, 45(5):1951–1987, 2017. doi: 10.1214/16-AOS1515.

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.

Donald B. Rubin. Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology*, 66(5):688–701, 1974. doi: 10.1037/h0037350.

Donald B. Rubin. Using multivariate matched sampling and regression adjustment to control bias in observational studies. *Journal of the American Statistical Association*, 74(366a):318–328, 1979. doi: 10.1080/01621459.1979.10482513.

Vira Semenova and Victor Chernozhukov. Debiased machine learning of conditional average treatment effects and other causal functions. *The Econometrics Journal*, 24(2):264–289, 2021. doi: 10.1093/ectj/utaa027.

Robert J. Serfling. *Approximation Theorems of Mathematical Statistics*. John Wiley & Sons, New York, 1980. ISBN 9780471024033. doi: 10.1002/9780470316481.

Aleksandr F. Timan. *Theory of Approximation of Functions of a Real Variable*. Pergamon Press, Oxford, 1963.

Anastasios A. Tsiatis. *Semiparametric Theory and Missing Data*. Springer, New York, 2006. ISBN 9780387324487.

Alexandre B. Tsybakov. *Introduction to Nonparametric Estimation*. Springer, New York, 2009. ISBN 9780387790510. doi: 10.1007/978-0-387-79052-7.

Gregory Valiant and Paul Valiant. Estimating the unseen. In *Proceedings of the Forty-Third Annual ACM Symposium on Theory of Computing*, STOC '11, pages 685–694. Association for Computing Machinery, 2011a. doi: 10.1145/1993636.1993727.

Gregory Valiant and Paul Valiant. The power of linear estimators. In *Proceedings of the 2011 IEEE 52nd Annual Symposium on Foundations of Computer Science*, FOCS '11, pages 403–412. IEEE, 2011b. doi: 10.1109/FOCS.2011.81.

Gregory Valiant and Paul Valiant. Instance optimal learning of discrete distributions. In *Proceedings of the Forty-Eighth Annual ACM Symposium on Theory of Computing*, STOC '16, pages 142–155. Association for Computing Machinery, 2016. doi: 10.1145/2897518.2897641.

Gregory Valiant and Paul Valiant. Estimating the unseen. *Journal of the ACM*, 64(6):1–41, 2017. doi: 10.1145/3125643.

Mark J. van der Laan and James M. Robins. *Unified Methods for Censored Longitudinal Data and Causality*. Springer, New York, 2003. ISBN 9780387955568. doi: 10.1007/978-0-387-21700-0.

Aad W. van der Vaart. *Asymptotic Statistics*. Cambridge University Press, Cambridge, 1998. ISBN 9780521496032.

Stefan Wager and Susan Athey. Estimation and inference of heterogeneous treatment effects using random forests. *Journal of the American Statistical Association*, 113(523):1228–1242, 2018. doi: 10.1080/01621459.2017.1319839.

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.

Yihong Wu and Pengkun Yang. Chebyshev polynomials, moment matching, and optimal estimation of the unseen. *The Annals of Statistics*, 47(2):857–883, 2019. doi: 10.1214/17-AOS1665.

Steve Yadlowsky. Explaining practical differences between treatment effect estimators with high dimensional asymptotics, 2022.

Zhenghao Zeng, Sivaraman Balakrishnan, Yanjun Han, and Edward H. Kennedy. Causal inference with high-dimensional discrete covariates, 2024. URL https://arxiv.org/abs/2405.00118. Revised 2026, version 3.