Minimax inference for threshold modified treatment policies with continuous treatments

CausalSmith*

September 11, 2026

Abstract

Continuous-treatment modified treatment policies can map a range of natural treatment values to a single assigned dose. This paper studies the exact lower-threshold clamp, which sends treatments below a deterministic threshold δ_n, the policy threshold, to δ_n and leaves larger treatments unchanged. Under a fixed finite-stratum observed-data model with bounded outcomes, Hölder outcome regression with known exponent β and radius L, and polynomial lower-tail treatment-density thinning $c_- a^\kappa \le \pi_x(a) \le c_+ a^\kappa$ with known exponent κ, the post-policy law contains a moving atom. The target is the observed clamp mean $\theta_{\delta_n}(P)$, the sum of the retained mean above the threshold and an atom-regression product at the boundary. The paper establishes matching minimax absolute-risk and uniformly honest expected-length rates

$$r_n = n^{-1/2} + \delta_n^{\kappa+1} h_n^{\beta}, \qquad n h_n^{2\beta+1}(\delta_n + h_n)^\kappa \asymp 1,$$

attained by a split-sample total-Gram stabilized local-polynomial estimator and a bias-aware interval. The phase diagram identifies the regular, critical, atom-dominated, fixed-threshold, and zero-threshold regimes induced by δ_n. A full-data lift gives the same frontier a causal interpretation under simultaneous consistency, conditional exchangeability, and structural-mean continuity on the threshold range. A continuity-only model yields the separate frontier $n^{-1/2} + \delta_n^{\kappa+1}$, with an explicit elbow and fixed-positive-threshold behavior.

1 Introduction

Modified treatment policies define interventions by transforming the treatment value that a unit would naturally receive. This idea is central in causal analyses where feasible interventions depend on the observed treatment process, including stochastic and modified policies for continuous or longitudinal treatments [Muñoz and van der Laan, 2011, Haneuse and Rotnitzky, 2013, Young et al., 2014, Díaz et al., 2021]. Applied implementations and tutorials make such policies operational, including threshold rules motivated by support and positivity considerations [Williams and Díaz, 2023, Hoffman et al., 2024]. This paper studies one exact rule: the lower-threshold clamp $d_\delta(a) = \max\{a, \delta\}$, which assigns all natural values below δ, the intervention threshold, to the boundary point δ.

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

The statistical feature of this policy is the atom created by the clamp. In a continuous-treatment experiment, values above δ remain on their natural continuous scale, while the lower tail collapses to a point mass at δ. The corresponding observed-data target in Definition 7 is

$$\theta_\delta(P) = \mathbb{E}_P\{Y\mathbf{1}(A > \delta)\} + \sum_{x\in\mathcal{X}} p_x q_x(\delta)\mu_x^P(\delta),$$

where $q_x(\delta)$, the lower-tail conditional treatment mass, is the probability collapsed by the clamp in stratum x, and $\mu_x^P(\delta)$, the threshold regression value, is the outcome regression at the boundary. This decomposition separates an ordinary retained-course mean from an atom-regression product. The latter term is the source of the boundary inference problem.

The observed model in Definition 5 fixes a finite stratum space, bounded outcomes, a Hölder regression class with known smoothness exponent β and radius L, and a polynomial treatment-density envelope

$$c_- a^\kappa \le \pi_x(a) \le c_+ a^\kappa$$

near the lower endpoint, with known thinning exponent κ. The deterministic threshold sequence δ_n may move with sample size. The information-balance bandwidth h_n in Definition 8 is defined by the local balance

$$n h_n^{2\beta+1}(\delta_n + h_n)^\kappa \asymp 1.$$

This balance combines the usual local-polynomial bias scale h_n^β with the effective local design mass under polynomial thinning. Since the clamp weights the boundary regression by an atom of order $\delta_n^{\kappa+1}$, the resulting frontier rate in Definition 9 is

$$r_n = n^{-1/2} + \delta_n^{\kappa+1} h_n^\beta.$$

The first main result is the minimax clamp frontier. Theorem 1 establishes matching upper and lower bounds for absolute-error risk over the observed Hölder model $\mathcal{M}$. The upper bound is attained by the total-Gram stabilized estimator of Algorithm 1, which estimates the retained mean, the lower-tail atom masses, and the boundary regression values on separate deterministic split blocks. The local-polynomial component is used on realized designs with sufficient total Gram curvature and is paired with a bounded fallback on singular designs. The lower bound combines two same-class experiments: a global Bernoulli mean shift for the $n^{-1/2}$ term and a localized Bernoulli perturbation at the moving threshold for the atom-weighted term.

The second main result gives honest inference at the same scale. Theorem 2 proves that the bias-aware interval in Algorithm 2 has uniform coverage over $\mathcal{M}$ and worst-case expected length of order r_n, and that every uniformly honest observed-sample interval has worst-case expected length at least a constant multiple of r_n. The interval uses bounded-outcome concentration for the retained mean and atom-mass estimates, and a Hölder bias plus empirical weight-energy radius for the threshold regression. This connects the clamp problem to honest nonparametric confidence-interval theory [Low, 1997, Armstrong and Kolesár, 2018, Armstrong and Kolesár, 2016] while preserving the atom-weighted structure of the target.

The phase diagram in Theorem 3 expands the compact rate r_n. Two deterministic scales organize the regimes: the risk-transition scale

$$\delta_{\mathrm{crit},n} = n^{-1/\{2(\beta\kappa+2\beta+\kappa+1)\}},$$

and the bandwidth-transition scale

$$\delta_{\mathrm{edge},n} = n^{-1/(2\beta+\kappa+1)}.$$

Thresholds with $\delta_n/\delta_{\mathrm{crit},n} \to 0$, or with $\delta_n/\delta_{\mathrm{crit},n} \to c_0 \in (0, \infty)$, have root-$n$ order. Vanishing thresholds above the critical scale enter an atom-dominated regime with rate

$$\delta_n^{\kappa+1}(n\delta_n^{\kappa})^{-\beta/(2\beta+1)}.$$

Thresholds converging to a fixed positive value have the ordinary positive-density pointwise Hölder exponent $\beta/(2\beta+1)$, weighted by an atom mass bounded away from zero. At threshold zero, the atom vanishes and the target reduces to the observed mean.

The causal interpretation is obtained through a full-data class tailored to the same lower-threshold policy. Members of $\mathcal{M}^{\mathrm{F}}$ in Definition 20 have observed margins in $\mathcal{M}$ and satisfy simultaneous structural consistency, conditional exchangeability, and continuity of the structural response mean on the threshold range. Proposition 2 identifies the causal clamp mean $\psi_\delta(P^{\mathrm{F}})$ with the observed target $\theta_\delta(P)$. Proposition 3 shows that the observed-margin map covers $\mathcal{M}$ and aligns the decision criteria. Consequently, Theorem 4 transports the same minimax risk and honest-length frontier to the causal clamp mean on $\mathcal{M}^{\mathrm{F}}$.

The paper also studies a continuity-only counterpart. In Definition 24, the density, stratum-mass, and polynomial-thinning restrictions are retained, while the regression condition supplies a memberwise continuous extension on the threshold interval. Theorem 5 characterizes the resulting frontier as

$$s_n = n^{-1/2} + \delta_n^{\kappa+1}.$$

The fallback estimator and Hoeffding interval attain this scale, and the minimax risk and honest expected length are bounded below at the same order. The elbow occurs at $\delta_n \asymp n^{-1/[2(\kappa+1)]}$, where the collapsed atom mass reaches the sampling scale. For fixed positive thresholds, the theorem records persistent minimax risk and honest length under qualitative continuity.

The results sit at the intersection of continuous-treatment causal inference, degenerate-design regression, and honest nonparametric inference. Existing work on continuous-treatment dose-response and modified-policy estimands develops identification, efficient estimation, targeted learning, threshold-response functionals, stochastic shifts, and support-aware policy analysis under their corresponding regularity structures [Imbens, 2000, Hirano and Imbens, 2005, Imai and van Dyk, 2004, Kennedy et al., 2017, van der Laan et al., 2022, Díaz et al., 2026, Koo et al., 2026, McCoy et al., 2026]. Degenerate-design regression theory gives pointwise rates when the design density vanishes near the target [Gaïffas, 2005, 2007], and classical minimax and local-polynomial theory provide the positive-density benchmark [Stone, 1982, Fan and Gijbels, 1996, Tsybakov, 2009]. The contribution here is the threshold-atom problem generated by the exact deterministic clamp: the local regression value is learned under thinning and then multiplied by the policy-induced atom mass.

Standing conditions. One set of conditions is in force throughout and is not repeated in the discussion surrounding each statement. The class constants $J, \beta, L, \kappa, c_-, c_+, p_{\min}, \bar{\delta}, \alpha$, the deterministic threshold path (δ_n), and the deterministic sample-split blocks are fixed before the sample is drawn and are available to every procedure; observations are independent and identically distributed from the observed law; the local-polynomial degree is the one induced by the smoothness exponent; and estimators and interval endpoints take values in the outcome range $[0, 1]$. All uniformity statements range over laws sharing these constants, and all risks, coverage probabilities, and expected lengths are computed under the n-fold product law of the observed data. The formal statements repeat these conditions in their hypotheses because each is separately checked, but a reader may treat them as one standing regime-and-sampling condition.

Each model, target, estimator, interval, and criterion has exactly one canonical definition: the model class is Definition 5, the target is Definition 7, and the estimator and interval are Algorithms 1 and 2. A few later environments introduce split-indexed aliases of those objects, written with an explicit block subscript, so that the finite-sample criteria can name a particular split; each such environment says which canonical definition it specializes and adds no new content. Section A is the verification supplement: it holds the auxiliary design, coercivity, and local-polynomial algebra on which the main proofs rest, and nothing in it is needed to read the main results.

Selected formal statements are proved in the verification supplement, with the precise verification scope and reproduction instructions given in Section A.

The rest of the paper proceeds as follows. Section 2 positions the paper within modified-policy causal inference, continuous-treatment estimation, degenerate-design regression, and honest confidence intervals. Section 3 defines the observed model, clamp policy, target, bandwidth, and frontier rate. Section 4 constructs the total-Gram estimator and bias-aware interval. Section 5 states the minimax risk, honest-length, phase-diagram, and calibration results. Section 6 gives the full-data causal lift. Section 7 treats the continuity-only model. Section 8 discusses scope, limitations, and open questions, including exact-modulus calibration.

2 Related work

Structural and longitudinal causal frameworks provide the counterfactual language for interventions and treatment regimes [Pearl, 2009, Robins et al., 2000, van der Laan and Robins, 2003, van der Laan and Rose, 2011]. Within that language, theoretical work develops identification and estimation for stochastic and modified policies in continuous-treatment settings, including rules whose assigned dose depends on the observed natural value [Muñoz and van der Laan, 2011, Haneuse and Rotnitzky, 2013, Young et al., 2014, Díaz et al., 2021]. Software and applied guidance make these policies operational and illustrate threshold interventions motivated by positivity and support constraints [Williams and Díaz, 2023, Hoffman et al., 2024]. The present paper studies the observed-data experiment generated by a lower-threshold modified treatment policy in which the deterministic rule clamps natural treatment values below the threshold to the threshold itself. That exact clamp creates a point mass at the boundary of the post-policy distribution, so the estimand combines an ordinary retained natural-course component with a boundary regression value multiplied by the mass collapsed by the policy.

A second line of work studies continuous-treatment dose-response functionals and the role of overlap in semiparametric and nonparametric inference [Imbens, 2000, Hirano and Imbens, 2005, Imai and van Dyk, 2004, Kennedy et al., 2017, Bonvini and Kennedy, 2026]. Recent natural-value-policy work develops root-n targeted-learning inference under its regularity and nuisance-rate conditions [Díaz et al., 2026], while positivity-violation work separates an in-support point-identified component from a Lipschitz-bounded extrapolation component [Koo et al., 2026]. Threshold-response methods target support-restricted stochastic functionals with efficient estimation and simultaneous bands [van der Laan et al., 2022], and stochastic-shift work targets effect modification with valid confidence intervals [McCoy et al., 2026]. The present estimand has a different post-policy law: its deterministic clamp maps the entire lower tail to one moving boundary point, producing a Dirac component whose weight and regression value must be learned together. Under a global two-sided polynomial envelope, the delivered results characterize minimax risk and uniformly honest expected length for this unsmoothed clamp functional.

The nonparametric component is closely related to regression with degenerate design. For smoothness s and a design density regularly varying with exponent γ, Gaïffas [2005] obtains the pointwise rate $n^{-s/(1+2s+\gamma)}$, up to slowly varying factors; the classical positive-density Hölder rate $n^{-s/(2s+1)}$ from Stone [1982] and the local-polynomial constructions surveyed by Fan and Gijbels [1996] provide the interior reference. In the present boundary regime, the information balance $nh_n^{2\beta+1}(\delta_n + h_n)^\kappa \asymp 1$ recovers the degenerate-design bandwidth when δ_n is at the edge scale. Farther into the interior it gives $h_n \asymp (n\delta_n^\kappa)^{-1/(2\beta+1)}$. The clamp target then multiplies the resulting boundary-regression error by the collapsed mass of order $\delta_n^{\kappa+1}$, yielding the atom term $\delta_n^{\kappa+1}h_n^\beta$ alongside the retained-mean $n^{-1/2}$ term. This mass weighting, absent from ordinary pointwise regression, generates the critical and atom-dominated threshold regimes. Two antecedents deserve a closer comparison, because they bear directly on the oracle conditions used here and on the sharp-constant question left open below.

The first is adaptation under degenerate design. Gaïffas [2007] studies pointwise estimation from inhomogeneous data when the smoothness is unknown, and obtains the adaptive rate with the usual logarithmic payment: the exponent matches the known-smoothness degenerate-design rate with n replaced by $n/\log n$, so adaptation costs a logarithmic factor rather than a change of exponent. The procedures in the present paper are class-indexed in the oracle sense: the smoothness exponent and radius, the thinning exponent and envelope constants, and the threshold path are fixed before sampling. Atom weighting changes the adaptation question rather than inheriting it. The clamp target multiplies the boundary-regression error by the collapsed mass $\delta_n^{\kappa+1}$, so a logarithmic loss in the regression component is invisible whenever the retained-mean term $n^{-1/2}$ dominates, and is felt only above the critical threshold scale; moreover the mass itself depends on the thinning exponent, so an adaptive treatment would have to adapt to κ as well as to β. The frontier reported here is the known-class benchmark for that question: it holds with β, L, and κ fixed and available to the procedure, which is the reference an adaptive treatment would be measured against. Section 8 takes up adaptation to unknown β, L, and κ as a question for future work.

The second is the modulus of continuity. The classical account of minimax rates through a modulus is due to Donoho and Liu [1991b], who show that the difficulty of estimating a linear functional over a convex class is governed by the between-class modulus of the functional with respect to a statistical distance, and Cai and Low [2004b] develop the corresponding theory for confidence intervals, where expected length over a subclass is governed by a within-class or between-class modulus rather than by the minimax rate alone. The honest expected-length lower bound established here is proved by direct two-point and localized-perturbation arguments, and it matches the risk frontier in order. Relating that frontier to a modulus is a well-posed next step, and the structure of the clamp target says what such a modulus would have to accommodate: the functional is linear in the outcome regression, while its weight $q_x(\delta)$ is determined by the design, so the appropriate family is a design-indexed one rather than the convex class of a single fixed linear functional. The sharp-constant question stated in the discussion is exactly this calibration question.

The inference results also build on the theory of honest confidence intervals. Low [1997] shows how expected length lower bounds constrain honest adaptation over nonparametric classes, and Armstrong and Kolesár [2018], Armstrong and Kolesár [2016] develop bias-aware interval methods that balance stochastic error and worst-case approximation bias. Those ideas are especially natural here because the clamp target contains a boundary regression term whose bias is controlled by smoothness and whose variance is controlled by the local design mass. The paper establishes matched minimax absolute-risk and uniformly honest expected-length rates for this target, so the estimation and interval scales agree under the same polynomial-thinning conditions.

Finally, the deterministic lower-threshold clamp studied here has a mixed-measure post-intervention law: values already above the threshold remain on their continuous natural scale, while values below the threshold collapse to the boundary point. That perspective explains why the lower-threshold clamp differs from smooth dose-response estimation: the intervention produces an atom whose mass is estimable from treatment frequencies, while its contribution to the mean depends on a boundary response value. The analysis below isolates this atom-regression product and derives the rates for both the Hölder model and the continuity-only model, thereby connecting modified-policy causal estimands to degenerate-design regression and honest nonparametric inference.

3 Setup and assumptions

We work with deterministic sequences and fixed constants throughout. Generic finite constants may change from line to line and depend only on the displayed model parameters. All asymptotic notation is along the sample size n; indices x range over the finite stratum space $\mathcal{X}$, and all probabilities and expectations are taken under the observed law P unless a different law is displayed. The observed unit is $O = (X, A, Y)$, with baseline stratum X, treatment A, and outcome Y. The deterministic regime constants include the smoothness exponent β, Hölder radius L, thinning exponent κ, density constants c_- and c_+, stratum-mass floor $p_{\min}$, threshold envelope $\bar{\delta}$, noncoverage level α, and deterministic threshold sequence δ_n.

The procedures below use fixed sample splitting and product sampling notation. The split is deterministic, so the probability statements condition on no random partitioning device.

Definition 1. [`synth_8`] (Admissible three-way split) *At sample size n, an admissible three-way split is a deterministic triple $B_n = (I_0, I_1, I_2)$ of pairwise-disjoint subsets of $\{1, \dots, n\}$ such that $|I_j| \geq \lfloor n/4 \rfloor$ for $j = 0, 1, 2$. The blocks are fixed before observing the sample.*

Definition 1 fixes the block structure used by the estimator and interval. The lower bounds on the block sizes keep each empirical component on the same sample-size scale while leaving the blocks independent under product sampling.

Definition 2. [`synth_16`] (Independent product sample laws) *For an observed-data law P on $\mathcal{X} \times [0, 1] \times [0, 1]$, write $P^{\otimes n}$ for the n-fold product law governing the observed sample $(O_1, \dots, O_n)$, under which the coordinates are independent with common law P. For a full-data law package P^{F}, write $(P^{\mathrm{F}})^{\otimes n}$ for the n-fold product law governing n independent full-data draws from P^{F}; estimators and intervals indexed by this law are measurable functions of the induced observed coordinates $(O_1, \dots, O_n)$.*

The notation in Definition 2 is used for risk and coverage statements in the observed and causal formulations. It records that full-data sampling still feeds estimators through the observed coordinates.

The sampling statement fixes the i.i.d. observed-data experiment used by the estimators and risk criteria.

Assumption 1. [ass:iid-sampling] (I.i.d. sampling) *The observed units $O_1, \dots, O_n$ are independent and identically distributed according to P.*

The independent random-design sampling condition in Assumption 1 is the standard sampling structure for the boundary-regression experiment [Gaïffas, 2005]. It lets the mass above the threshold, the empirical lower-threshold atom, and the local regression fit be analyzed with product-law concentration under the same observed law.

The next conditions describe the treatment distribution within strata and the regression function at the boundary created by the lower-threshold clamp.

Assumption 2. [ass:conditional-density-law] (Conditional treatment density) *For each $x \in \mathcal{X}$, the map $a \mapsto \pi_x(a)$ is Borel measurable on $[0,1]$, and $\pi_x(a) \geq 0$ for Lebesgue-a.e. $a \in [0,1]$. Moreover, $p_x = P(X = x)$, and for every Borel set $B \subseteq [0,1]$,*

$$P(X = x,\ A \in B) = p_x \int_B \pi_x(a)\,\mathrm{d}a.$$

The conditional absolute-continuity requirement in Assumption 2 is standard for continuous-treatment modified policies [Díaz et al., 2021]. It introduces the stratum density π_x and the stratum mass p_x, so the probability collapsed by a threshold can be expressed as an ordinary integral.

Assumption 3. [ass:stratum-mass] (Stratum mass lower bound) *For every $x \in \mathcal{X}$,*

$$p_x \geq p_{\min}.$$

The finite-stratum overlap condition in Assumption 3 is standard in stratified continuous-treatment analyses [Díaz et al., 2021]. It ensures that every stratum contributes a nonvanishing share of the sample, which keeps the local design problem from being driven by empty baseline cells.

Assumption 4. [ass:polynomial-thinning] (Polynomial overlap thinning) *For every $x \in \mathcal{X}$ and for Lebesgue-almost every $a \in [0,1]$,*

$$c_- a^\kappa \leq \pi_x(a) \leq c_+ a^\kappa.$$

Assumption 4 is specific to this analysis. It states that treatment support near the lower boundary thins at a polynomial order, so the amount of information available to learn the boundary regression value is governed by the same exponent in every stratum. The two sides of the envelope do different work, and it helps to keep them apart. The lower bound $c_- a^\kappa \leq \pi_x(a)$ guarantees local information: it lower bounds the expected count and the local Gram eigenvalue in the threshold window, and it lower bounds the collapsed atom mass $q_x(\delta)$. It is therefore the side used for Gram coercivity and for atom-mass lower bounds. The upper bound $\pi_x(a) \leq c_+ a^\kappa$ controls upper-risk quantities: the same window mass, the local Gram entries, and the collapsed atom mass entering the bias and length calculations. The lower-bound experiments use the canonical density $\pi_x(a) = (\kappa + 1) a^\kappa$, which the regime constraint $c_- \leq \kappa + 1 \leq c_+$ admits, and the likelihood divergences are computed for that canonical density and its localized perturbations. Densities proportional to a^κ, and bounded multiplicative perturbations $a^\kappa r_x(a)$ with $0 < c_- \leq r_x(a) \leq c_+$, are representative members.

Definition 3. [synth_19] (Local-polynomial degree) *For each smoothness exponent $\beta \in \mathbb{R}$, the local-polynomial degree is*

$$\ell(\beta) := \max\{\lceil \beta \rceil_{\mathbb{N}} - 1, 0\},$$

where $\lceil \beta \rceil_{\mathbb{N}}$ denotes the least natural number greater than or equal to β.

The degree ℓ in Definition 3 is the order used by the local polynomial at the threshold. It matches the Taylor-remainder smoothness imposed next.

Assumption 5. [ass:holder-regression] (Hölder regression regularity) *The regression function μ_x^P satisfies the following condition with parameters β and L. For every $x \in \mathcal{X}$:*

- ***(Continuity.)*** *The map $a \mapsto \mu_x^P(a)$ is continuous on $[0, 1]$.*
- ***(Range.)*** *For every $a \in [0, 1]$, $\mu_x^P(a) \in [0, 1]$.*
- ***(Regression version.)*** *The function $\mu_X^P(A)$ realizes the conditional outcome regression given the complete design:*

$$\mathbb{E}_P[Y \mid X, A] = \mu_X^P(A) \quad P\text{-almost surely.}$$

- ***(Hölder Taylor remainder.)*** *Let $\ell = \ell(\beta)$ be the local-polynomial order defined in Definition 3. For all $s, t \in [0, 1]$, with derivatives taken intrinsically on $[0, 1]$,*

$$\left| \mu_x^P(t) - \sum_{j=0}^{\ell} \frac{(\mu_x^P)^{(j)}(s)}{j!}(t-s)^j \right| \le L|t-s|^\beta.$$

The Taylor-remainder formulation in Assumption 5 is a standard Hölder outcome-regression smoothness condition for bias-aware nonparametric inference [Armstrong and Kolesár, 2016]. In this model, the boundary value of the regression μ_x^P is multiplied by the threshold mass, so the condition supplies the deterministic approximation control used by the local-polynomial estimator.

These ingredients define the observed model class and the lower-threshold clamp target.

Definition 4. [synth_18] (Observed continuity class $\mathcal{M}_{\text{cont}}$) *Let $\mathcal{X} = \{1, \ldots, J\}$, let $O = (X, A, Y) \in \mathcal{X} \times [0, 1] \times [0, 1]$, and let $P \in \mathcal{P}(\mathcal{X} \times [0, 1] \times [0, 1])$. Write*

$$p_x = P(X = x).$$

The law P belongs to the continuity-only class $\mathcal{M}_{\text{cont}}$ when, for every $x \in \mathcal{X}$, the conditional-density condition of Assumption 2, the stratum-mass condition of Assumption 3, and the polynomial-thinning condition of Assumption 4 hold, and an almost-everywhere version of $a \mapsto \mathbb{E}_P[Y \mid A = a, X = x]$ has a unique continuous extension

$$\mu_{x,\text{cont}}^P : [0, \bar{\delta}] \to [0, 1].$$

The Hölder class $\mathcal{M}$ imposes the same three design conditions together with the Hölder regression condition of Assumption 5; its canonical definition is Definition 5, and the canonical continuity-only definition is Definition 24.

Definition 4 records the continuity-only class $\mathcal{M}_{\rm cont}$ alongside the Hölder class. The later continuity-only section uses the selected continuous regression extension $\mu^P_{x,\rm cont}$ when smoothness is replaced by qualitative continuity on the threshold envelope.

Definition 5. [def:model-class] (Observed law class $\mathcal{M}$) *For $P \in \mathcal{P}(\mathcal{X} \times [0,1] \times [0,1])$, let*

$$p_x = P(X = x), \qquad x \in \mathcal{X}.$$

The observed-data law class $\mathcal{M}$ is

$$\mathcal{M} = \{P \in \mathcal{P}(\mathcal{X} \times [0,1] \times [0,1]) : P \text{ satisfies the conditions listed below}\}.$$

Membership means that P satisfies:

- ***(Conditional density.)*** *The conditional-density condition in Assumption 2.*
- ***(Stratum mass.)*** *The stratum-mass condition in Assumption 3.*
- ***(Polynomial thinning.)*** *The polynomial-thinning condition in Assumption 4.*
- ***(Hölder regression.)*** *The Hölder member condition in Assumption 5.*

The observed class $\mathcal{M}$ in Definition 5 collects the design and regression restrictions under one membership statement. This compact notation is the domain for the estimator, honest interval, and minimax criteria in the next two sections.

Definition 6. [def:clamp-policy] (Clamp policy d_δ and A^δ) *For $0 \le \delta \le \bar{\delta}$, the lower-threshold clamp map d_δ and the clamped treatment A^δ are*

$$d_\delta(a) = \max\{a, \delta\}, \qquad A^\delta = d_\delta(A).$$

The map d_δ in Definition 6 leaves treatments above the threshold unchanged and moves lower treatments to the threshold. The resulting target therefore combines a natural-course component above δ with a boundary-regression component at δ.

Definition 7. [def:clamp-functional] (Clamp functionals $q_x(\delta)$, $\nu_\delta(P)$, and $\theta_\delta(P)$) *For $P \in \mathcal{M}$ as in Definition 5, $x \in \mathcal{X}$, and $0 \le \delta \le \bar{\delta}$, define*

$$q_x(\delta) = \int_{[0,\delta]} \pi_x(a)\,\mathrm{d}a, \qquad \nu_\delta(P) = \mathbb{E}_P[Y\mathbf{1}\{A > \delta\}].$$

The unsmoothed observed-data clamp target is

$$\theta_\delta(P) = \nu_\delta(P) + \sum_{x\in\mathcal{X}} p_x q_x(\delta)\mu^P_x(\delta).$$

In Definition 7, $q_x(\delta)$ is the conditional probability mass collapsed by the clamp in stratum x, $\nu_\delta(P)$ is the retained mean above the threshold, and $\theta_\delta(P)$ is the observed-data clamp target. The expression isolates the atom-regression product that drives the boundary inference problem.

At sample size n the target is written $\theta_{\delta_n}(P)$: it is the functional of Definition 7 evaluated at the deterministic threshold $\delta = \delta_n$.

The remaining definitions set the bandwidth, rate scale, and range projection used by the estimation and interval statements.

Definition 8. [`def:bandwidth`] (Information-balance bandwidth h_n) *The information-balance bandwidth h_n for the threshold sequence δ_n is*

$$h_n = \inf\left\{h \in (0, 1-\bar{\delta}] : nh^{2\beta+1}(\delta_n + h)^{\kappa} \geq 1\right\},$$

with $h_n = 1 - \bar{\delta}$ when the displayed set is empty.

The bandwidth h_n in Definition 8 is chosen by the local information available in a window to the right of the threshold. The factor $(\delta_n + h)^{\kappa}$ carries the polynomial thinning from Assumption 4 into the window-mass calculation.

Definition 9. [`def:frontier`] (Frontier rate r_n) *The Hölder minimax rate r_n is*

$$r_n = n^{-1/2} + \delta_n^{\kappa+1} h_n^{\beta}.$$

The rate r_n in Definition 9 combines the regular root-n contribution from the retained mean and threshold mass with the atom-weighted Hölder approximation scale. This is the benchmark used for the minimax risk and honest expected-length results.

Definition 10. [`synth_6`] (Projection onto the outcome range) *For a real number t, define $\Pi_{[0,1]}(t) = \min\{1, \max\{0, t\}\}$. Applied to an estimator or interval endpoint, $\Pi_{[0,1]}$ denotes this pointwise projection onto the outcome range $[0, 1]$.*

Definition 11. [`synth_2`] (Continuity-only regression extension) *For $P \in \mathcal{M}_{\mathrm{cont}}$ and $x \in \mathcal{X}$, $\mu^P_{x,\mathrm{cont}}$ denotes the unique continuous function on $[0, \bar{\delta}]$ that agrees, almost everywhere on that interval under the conditional treatment law given $X = x$, with a version of the conditional mean $a \mapsto \mathbb{E}_P[Y \mid A = a, X = x]$.*

Definition 10 records the range projection used later for estimators and confidence endpoints. Because the target is built from an outcome regression taking values in $[0, 1]$, the projection keeps the reported numerical object on the same outcome scale.

4 Estimation and honest intervals

The estimator separates the three empirical tasks in $\theta_{\delta_n}(P)$: estimating the retained mean above the threshold, estimating the mass collapsed by the clamp, and estimating the boundary regression value that multiplies that mass. The deterministic split from Definition 1 assigns one block to the retained-course mean, a second to the lower-threshold atom masses, and a third to the local-polynomial fit at the boundary. This separation lets the interval treat the three components with direct concentration bounds for bounded variables. The local regression step follows the boundary-polynomial logic of Fan and Gijbels [1996], with a stabilization rule tailored to the total Gram matrix in each stratum.

The realized-design notation is as follows. It records the rescaled local coordinate, the $v_\ell(u)$ monomial basis, the stratum-specific local count, and the event on which the total Gram matrix has enough curvature to support the intercept fit.

Definition 12. [`def:local-design-notation`] (Local design notation) *For a realized sample $z = (O_1, \ldots, O_n)$, deterministic split blocks I_0, I_1, I_2 as in Definition 1, a stratum $x \in \mathcal{X}$, local-polynomial degree ℓ, threshold δ_n, and bandwidth h_n, put*

$$U_i = \frac{A_i - \delta_n}{h_n}, \qquad v_\ell(u) = (1, u, \ldots, u^{\ell})^{\top}.$$

Define

$$N_x = \sum_{i \in I_2} \mathbf{1}\{X_i = x,\ A_i \in [\delta_n, \delta_n + h_n]\}$$

and

$$G_x = \sum_{i \in I_2} \mathbf{1}\{X_i = x,\ 0 \le U_i \le 1\} v_\ell(U_i) v_\ell(U_i)^\top.$$

For $\rho \ge 0$, define

$$M_{\ell,\kappa,\rho} = \frac{\int_0^1 v_\ell(u) v_\ell(u)^\top (\rho + u)^\kappa \, \mathrm{d}u}{\int_0^1 (\rho + u)^\kappa \, \mathrm{d}u},$$

and set

$$\lambda_\star = \frac{c_-}{c_+} \inf_{\rho \ge 0} \lambda_{\min}(M_{\ell,\kappa,\rho}) \, .$$

The good-design event Ω_x is the event that $N_x > 0$ and, for every $v \in \mathbb{R}^{\ell+1}$,

$$v^\top G_x v \ge \frac{\lambda_\star N_x}{2} \|v\|_2^2.$$

The exact local-polynomial intercept weight is

$$w_{ix} = \begin{cases} \mathbf{1}\{i \in I_2,\ X_i = x,\ 0 \le U_i \le 1\} \, e_0^\top G_x^{-1} v_\ell(U_i), & \textit{on } \Omega_x, \\ 0, & \textit{on } \Omega_x^c. \end{cases}$$

In Definition 12, the local count N_x and the total localized Gram matrix G_x are computed on the third split block I_2. The normalized population moment matrix $M_{\ell,\kappa,\rho}$ supplies the reference curvature induced by the polynomial thinning law in Assumption 4. The constant $\lambda_\star$ converts that reference curvature into a uniform empirical design threshold, and the good-design event Ω_x is the realized condition under which the intercept weights w_{ix} are used.

With these quantities fixed, the point estimator is a plug-in estimator for the decomposition in Definition 7. It estimates each summand on its own split block and applies the range projection introduced in Definition 10.

Algorithm 1 (Total-Gram estimator $\widehat{\theta}_n$)

Given split blocks I_0, I_1, I_2 and the local-design quantities in Definition 12, define the estimator as follows.

1. Compute the retained-course empirical mean

$$\widehat{\nu}_n = \frac{1}{|I_0|} \sum_{i \in I_0} Y_i \mathbf{1}\{A_i > \delta_n\}.$$

2. For each $x \in \mathcal{X}$, compute the empirical lower-threshold atom mass

$$\widehat{b}_x = \frac{1}{|I_1|} \sum_{i \in I_1} \mathbf{1}\{X_i = x,\ A_i \le \delta_n\}.$$

3. For each $x \in \mathcal{X}$, compute the clipped local-polynomial threshold estimate

$$\widetilde{\mu}_x = \begin{cases} \Pi_{[0,1]}\left(\sum_{i \in I_2} w_{ix} Y_i\right), & \Omega_x, \\ 1/2, & \Omega_x^c. \end{cases}$$

4. Output the total-Gram clamp-target estimator

$$\widehat{\theta}_n = \Pi_{[0,1]}\left[\widehat{\nu}_n + \sum_{x \in \mathcal{X}} \widehat{b}_x \widetilde{\mu}_x\right].$$

The estimator $\widehat{\theta}_n$ in Algorithm 1 uses the local-polynomial intercept only on Ω_x. On the complementary realized designs, the fallback value 1/2 keeps the atom-regression factor inside the bounded outcome range. The projection in the last line keeps the estimator on the same scale as $\theta_{\delta_n}(P)$.

For statements indexed by a full deterministic split B_n, it is useful to record the same construction with the split made explicit.

Algorithm 2 (Honest interval CI_n)

Given the local-design quantities in Definition 12, define the interval as follows.

1. Set the Hoeffding calibration multiplier and split-block stochastic-error radii

$$t_\alpha = \sqrt{\frac{\log(12J/\alpha)}{2}}, \qquad b_0 = \frac{t_\alpha}{\sqrt{|I_0|}}, \qquad b_1 = \frac{t_\alpha}{\sqrt{|I_1|}}.$$

2. For each $x \in \mathcal{X}$, set the local regression bias-plus-noise radius

$$B_x = L h_n^\beta \sum_{i \in I_2} |w_{ix}| + t_\alpha \left(\sum_{i \in I_2} w_{ix}^2\right)^{1/2}.$$

3. For each $x \in \mathcal{X}$, set the atom-weighted interval contribution

$$D_x = \begin{cases} (\widehat{b}_x + b_1) B_x + b_1, & \Omega_x, \\ \widehat{b}_x + b_1, & \Omega_x^c. \end{cases}$$

4. Output the bias-aware honest interval

$$\mathrm{CI}_n = \left[\widehat{\theta}_n - b_0 - \sum_{x \in \mathcal{X}} D_x,\ \widehat{\theta}_n + b_0 + \sum_{x \in \mathcal{X}} D_x\right] \cap [0,1].$$

Definition 13 is the version used by the finite-sample risk and length criteria below. It makes explicit that the sampling split is fixed before the data are observed, while the good-design events and intercept weights are realized from the third block.

The confidence interval adds deterministic bias allowances to concentration radii. The stochastic terms use Hoeffding-type calibration for bounded outcomes and indicators [Hoeffding, 1963], while the local regression radius combines a Hölder bias bound with the empirical ℓ_2 size of the intercept weights. This bias-aware construction is in the spirit of honest nonparametric inference with smoothness restrictions [Armstrong and Kolesár, 2018, Armstrong and Kolesár, 2016].

Definition 13. [`synth_4`] (Total-Gram stabilized estimator) *Let $B_n = (I_{0,n}, I_{1,n}, I_{2,n})$ be a deterministic admissible three-way split, let h_n be the information-balance bandwidth, and let ℓ be the local-polynomial order. Define the projection*

$$\operatorname{clamp}_{[0,1]}(t) = \min\{1, \max\{0, t\}\}.$$

Set

$$\widehat{\nu}_n = |I_{0,n}|^{-1} \sum_{i \in I_{0,n}} Y_i \mathbf{1}\{A_i > \delta_n\}, \qquad \widehat{b}_x = |I_{1,n}|^{-1} \sum_{i \in I_{1,n}} \mathbf{1}\{X_i = x, A_i \le \delta_n\}.$$

For each x, let $\widetilde{\mu}_x = \operatorname{clamp}_{[0,1]}(\sum_{i \in I_{2,n}} w_{ix} Y_i)$ on Ω_x, and let $\widetilde{\mu}_x = 1/2$ on Ω_x^c. The split-indexed total-Gram estimator is

$$\widehat{\theta}^{\mathrm{TG}}_{n,B_n} = \operatorname{clamp}_{[0,1]}\left(\widehat{\nu}_n + \sum_{x \in \mathcal{X}} \widehat{b}_x \widetilde{\mu}_x \right).$$

Here J, the number of strata, enters only through the Hoeffding multiplier t_α. The interval CI_n in Algorithm 2 treats the retained mean, atom masses, and local-regression intercepts as separate empirical components. On Ω_x, D_x carries both the uncertainty in $\widehat{b}_x$ and the bias-plus-noise radius for the threshold regression. On Ω_x^c, the contribution covers the full feasible atom-regression range, which is the branch used to preserve the same bounded-outcome scale on singular realized designs.

The split-indexed interval records the same accounting with B_n attached to each empirical quantity.

Definition 14. [`synth_12`] (Bias-aware interval for a fixed split B_n) *Let $B_n = (I_{0,n}, I_{1,n}, I_{2,n})$ be an admissible deterministic three-way split at sample size n. With h_n and δ_n fixed, form the total-Gram weights w_{ix} and good-design events Ω_x from the third block $I_{2,n}$. Set $\widehat{\nu}_{n,B_n} = |I_{0,n}|^{-1} \sum_{i \in I_{0,n}} Y_i \mathbf{1}\{A_i > \delta_n\}$, $\widehat{b}_{x,B_n} = |I_{1,n}|^{-1} \sum_{i \in I_{1,n}} \mathbf{1}\{X_i = x, A_i \le \delta_n\}$, and $\widetilde{\mu}_{x,B_n} = \Pi_{[0,1]}(\sum_{i \in I_{2,n}} w_{ix} Y_i)$ on Ω_x, with $\widetilde{\mu}_{x,B_n} = 1/2$ on Ω_x^c. Define $\widehat{\theta}^{\mathrm{TG}}_{n,B_n} = \Pi_{[0,1]}\{\widehat{\nu}_{n,B_n} + \sum_{x \in \mathcal{X}} \widehat{b}_{x,B_n} \widetilde{\mu}_{x,B_n}\}$. Let $t_\alpha = \sqrt{\log(12J/\alpha)/2}$, $b_{0,B_n} = t_\alpha / \sqrt{|I_{0,n}|}$, and $b_{1,B_n} = t_\alpha / \sqrt{|I_{1,n}|}$. For each $x \in \mathcal{X}$, set $B_{x,B_n} = L h_n^\beta \sum_{i \in I_{2,n}} |w_{ix}| + t_\alpha (\sum_{i \in I_{2,n}} w_{ix}^2)^{1/2}$ and $D_{x,B_n} = (\widehat{b}_{x,B_n} + b_{1,B_n}) B_{x,B_n} + b_{1,B_n}$ on Ω_x, while $D_{x,B_n} = \widehat{b}_{x,B_n} + b_{1,B_n}$ on Ω_x^c. The bias-aware interval formed from B_n is $\mathrm{CI}_{n,B_n} = [\widehat{\theta}^{\mathrm{TG}}_{n,B_n} - b_{0,B_n} - \sum_{x \in \mathcal{X}} D_{x,B_n}, \widehat{\theta}^{\mathrm{TG}}_{n,B_n} + b_{0,B_n} + \sum_{x \in \mathcal{X}} D_{x,B_n}] \cap [0,1]$.*

Definition 14 fixes the interval whose coverage and expected length are evaluated in the main statistical results. Its two branches mirror Algorithm 2: the regular branch combines atom-mass uncertainty with local-polynomial bias and noise, and the fallback branch uses the bounded outcome range when the realized design lacks sufficient Gram curvature.

The length criterion is pointwise in the observed sample and then averaged under the observed product law.

Definition 15. [synth_17] (Length of an observed-sample interval) *For an observed-sample measurable interval* $C_n = C_n(O_1, \dots, O_n) = [\ell_n(O_1, \dots, O_n), u_n(O_1, \dots, O_n)] \subseteq [0, 1]$*, define its length pointwise by* $\operatorname{len}(C_n) = u_n(O_1, \dots, O_n) - \ell_n(O_1, \dots, O_n)$*. Thus expected lengths such as* $\mathbb{E}_{P^{\otimes n}}[\operatorname{len}(C_n)]$ *are taken with respect to the n-fold observed-sample law.*

Finally, the finite-sample criteria place the split-indexed estimator and interval on the common worst-case scale used in the minimax comparisons.

Definition 16. [synth_5] (Concrete stabilized risk and length) *For a deterministic admissible split* B_n*, define the finite-sample worst-case risk of the total-Gram estimator by*

$$\operatorname{Risk}_n(B_n) = \sup_{P \in \mathcal{M}} \mathbb{E}_{P^{\otimes n}} \left| \widehat{\theta}^{\mathrm{TG}}_{n,B_n} - \theta_{\delta_n}(P) \right|.$$

For the bias-aware interval CI_{n,B_n} *formed from the same split, define its worst-case expected length by*

$$\operatorname{Length}_n(B_n) = \sup_{P \in \mathcal{M}} \mathbb{E}_{P^{\otimes n}}[\operatorname{len}(\mathrm{CI}_{n,B_n})].$$

Definition 16 evaluates the point estimator and interval over the observed-law class $\mathcal{M}$ from Definition 5. The risk uses absolute error for the clamp target $\theta_{\delta_n}(P)$, and the length functional applies Definition 15 to the bias-aware interval from Definition 14.

5 Main statistical results

Section 4 constructed the split-sample total-Gram estimator and its bias-aware interval for the observed clamp target in Definition 7. We now state the corresponding performance guarantees. The product sampling law in Definition 2 and the interval-length convention in Definition 15 put point estimators, confidence intervals, and minimax benchmarks on the same finite-sample scale.

The scalar calibration at the end of the section uses a concrete Bernoulli pair. We record it before the main bounds so that the same notation can be used in the rate check.

Definition 17. [synth_10] (One-cell Bernoulli regressions) *In the one-stratum calibration with* $\beta = \kappa = 1$ *and* $\pi(a) = 2a$*, define the null Bernoulli regression by* $m_0(a) = 1/2$*. For a threshold* δ_n *and bandwidth* $h_n > 0$*, define the localized alternative by* $m_{1,\delta_n,h_n}(a) = 1/2 + h_n \, \tau_{\delta_n,h_n}(a)$*, where* $\tau_{\delta,h}(a) = \max\{0, 1 - |a - \delta|/h\}$*.*

The triangular alternative in Definition 17 is centered at the moving threshold and has height proportional to the bandwidth. It gives a transparent version of the same local perturbation that drives the lower-bound component of the general theory.

The minimax benchmarks optimize over all observed-sample procedures. The first criterion measures worst-case absolute accuracy for estimating the unsmoothed clamp target, and the second measures worst-case expected length among intervals with uniform coverage.

Definition 18. [synth_3] (Observed minimax risk and honest length) *For the observed clamp target* $\theta_{\delta_n}(P)$ *over* $\mathcal{M}$*, the minimax absolute-error risk is*

$$R_n^\star = \inf_{\widehat{\theta}} \sup_{P \in \mathcal{M}} \mathbb{E}_{P^{\otimes n}} \left| \widehat{\theta}(O_1, \dots, O_n) - \theta_{\delta_n}(P) \right|,$$

where the infimum ranges over observed-sample measurable estimators taking values in $[0,1]$. *The minimax honest expected length is*

$$L_n^\star = \inf_{C_n:\, \inf_{P\in\mathcal{M}} P^{\otimes n}\{\theta_{\delta_n}(P)\in C_n\}\ge 1-\alpha} \ \sup_{P\in\mathcal{M}} \mathbb{E}_{P^{\otimes n}}[\mathrm{len}(C_n)],$$

where C_n *ranges over observed-sample measurable interval procedures with endpoints in* $[0,1]$.

Thus $R_n^\star$ is the best possible worst-case absolute accuracy for the observed clamp target, while $L_n^\star$ is the best possible worst-case length under uniform honesty. These criteria match the honesty framework used in nonparametric confidence intervals [Low, 1997, Armstrong and Kolesár, 2018].

The main risk result gives matching upper and lower bounds at the rate in Definition 9. The upper bound is attained by the stabilized total-Gram estimator from Algorithm 1; the lower bound is witnessed by both a global mean shift and a localized threshold perturbation.

Theorem 1. [`thm:minimax-risk`] (Minimax clamp frontier) *Fix* $J, \beta, \kappa, L, c_-, c_+, p_{\min}, \bar{\delta}, \alpha$ *satisfying the regime conditions:*

- ***(Regime.)*** $J \ge 1$, $\beta > 0$, $\kappa \ge 0$, $L > 0$, $c_- > 0$, $c_- \le \kappa + 1 \le c_+$, $p_{\min} > 0$, $p_{\min} \le 1/J$, $\bar{\delta} \in (0,1)$, *and* $\alpha \in (0, 1/2)$.

Then there exist constants $0 < c < C$ *and an amplitude* $a \in (0, 1/4]$, *depending only on the displayed regime constants, such that, for every deterministic threshold sequence* $(\delta_n)_{n\ge1}$ *with* $\delta_n \in [0, \bar{\delta}]$ *for every* n *and every deterministic sequence* $B_n = (I_{0,n}, I_{1,n}, I_{2,n})$ *of three pairwise-disjoint sample blocks with* $|I_{j,n}| \ge \lfloor n/4 \rfloor$ *for* $j = 0, 1, 2$, *for all sufficiently large* n, *with* h_n *the information-balance bandwidth from Definition 8 and*

$$r_n = n^{-1/2} + \delta_n^{\kappa+1} h_n^{\beta}$$

as in Definition 9, the observed minimax absolute-error risk $R_n^\star$ *over the clamp model* $\mathcal{M}$ *in Definition 5 and the worst-case i.i.d. risk over* $\mathcal{M}$ *of the stabilized estimator using* B_n *satisfy*

$$c r_n \le R_n^\star \le \sup_{P\in\mathcal{M}} \mathbb{E}_P\left[\left|\widehat{\theta}^{\mathrm{TG}}_{n,B_n} - \theta_{\delta_n}(P)\right|\right] \le C r_n.$$

Moreover, for all sufficiently large n, *the lower bound is witnessed inside the same clamp model by:*

- ***(Global shift.)*** *Laws* $P_0, P_1 \in \mathcal{M}$ *satisfying Assumption 1, with the same* (X, A)*-design distribution and the same* p_x *and* π_x *functions, having Bernoulli outcomes, and obeying*

  $$\mu_x^{P_1}(a) = \mu_x^{P_0}(a) + n^{-1/2} \quad \textit{for every } x \textit{ and } a \in [0,1],$$

  with well-posed product chi-squared divergence,

  $$P_1^{\otimes n} \ll P_0^{\otimes n}, \qquad \int \left(\frac{dP_1^{\otimes n}}{dP_0^{\otimes n}} - 1\right)^2 dP_0^{\otimes n} < \infty,$$

  and

  $$c n^{-1/2} \le |\theta_{\delta_n}(P_1) - \theta_{\delta_n}(P_0)|, \qquad \chi^2\big(P_1^{\otimes n}, P_0^{\otimes n}\big) \le C.$$

- ***(Localized perturbation.)*** *Laws $Q_0, Q_1 \in \mathcal{M}$ satisfying Assumption 1, with the same (X, A)-design distribution and the same p_x and π_x functions, having Bernoulli outcomes, and admitting a continuous bump $b : \mathbb{R} \to \mathbb{R}$ satisfying $b(0) = 1$, $b(u) \geq 0$ for every u, $b(u) = 0$ for $u \notin [-1, 1]$, and $|b(u)| \leq 1$ for $u \in [-1, 1]$, with regression shift*

  $$\mu_x^{Q_1}(t) - \mu_x^{Q_0}(t) = a\, h_n^{\beta} b\left(\frac{t - \delta_n}{h_n}\right)$$

  for every x and $t \in [0, 1]$. The product chi-squared divergence is well posed,

  $$Q_1^{\otimes n} \ll Q_0^{\otimes n}, \qquad \int \left(\frac{dQ_1^{\otimes n}}{dQ_0^{\otimes n}} - 1\right)^2 dQ_0^{\otimes n} < \infty,$$

  and

  $$c\, \delta_n^{\kappa+1} h_n^{\beta} \leq |\theta_{\delta_n}(Q_1) - \theta_{\delta_n}(Q_0)|\,, \qquad \chi^2\big(Q_1^{\otimes n}, Q_0^{\otimes n}\big) \leq C.$$

The two lower-bound constructions in Theorem 1 isolate the two terms in r_n. A global Bernoulli shift produces the parametric contribution, while a Hölder-sized bump at the threshold produces the atom-weighted local contribution. The estimator in Algorithm 1 attains this combined scale, so the observed clamp problem has the same order from the attainable and unavoidable sides.

The corresponding interval result uses the same rate scale. Its statement combines finite-sample coverage for the constructed interval with a minimax lower bound for every honest observed-sample interval.

Theorem 2. [`thm:honest-length`] (Honest length frontier) *There are constants $c, C \in (0, \infty)$, with $c < C$, for which the following holds.*

- ***(Regime constants.)*** *The number of strata J is a positive integer, $\beta > 0$, $\kappa \geq 0$, $L > 0$, $c_- > 0$, $c_- \leq \kappa + 1 \leq c_+$, $p_{\min} > 0$, $p_{\min} \leq 1/J$, $\bar{\delta} \in (0, 1)$, and $\alpha \in (0, 1/2)$.*

- ***(Threshold path.)*** *The deterministic thresholds satisfy $0 \leq \delta_n \leq \bar{\delta}$ for every n.*

- ***(Sample splits.)*** *For every n, $B_n = (I_{0,n}, I_{1,n}, I_{2,n})$ is a deterministic three-way split of the sample indices with pairwise disjoint blocks and $|I_{j,n}| \geq \lfloor n/4 \rfloor$ for $j = 0, 1, 2$.*

For all sufficiently large n, set h_n to be the information-balance bandwidth for δ_n from Definition 8, and set

$$r_n = n^{-1/2} + \delta_n^{\kappa+1} h_n^{\beta}.$$

Then the stabilized interval CI_n has worst-case coverage

$$\inf\Big\{P^{\otimes n}\{\theta_{\delta_n}(P) \in \mathrm{CI}_n(O_1, \ldots, O_n)\} : P \in \mathcal{M}\Big\} \geq 1 - \alpha,$$

and its worst-case expected length is comparable to r_n:

$$c\, r_n \leq \sup\Big\{\mathbb{E}_{P^{\otimes n}}\big[\mathrm{len}(\mathrm{CI}_n)\big] : P \in \mathcal{M}\Big\} \leq C\, r_n.$$

Moreover, the minimax honest expected length satisfies

$$c\, r_n \leq L_n^{\star},$$

where $L_n^\star$ is the infimum, over observed-sample confidence procedures with uniform coverage at least $1-\alpha$ over $\mathcal{M}$, of their worst-case expected length. Equivalently, every observed-sample confidence procedure C_n with

$$\inf\Big\{P^{\otimes n}\{\theta_{\delta_n}(P) \in C_n(O_1,\dots,O_n)\} : P \in \mathcal{M}\Big\} \geq 1-\alpha$$

has worst-case expected length at least $c\,r_n$:

$$c\,r_n \leq \sup\Big\{\mathbb{E}_{P^{\otimes n}}\big[\mathrm{len}(C_n)\big] : P \in \mathcal{M}\Big\}.$$

The interval in Theorem 2 is calibrated to the same two sources of difficulty as the estimator. The retained-course and atom-mass terms are controlled by bounded-outcome concentration, while the threshold regression term carries the local Hölder bias and the realized Gram behavior encoded in Algorithm 2. The lower bound shows that uniform honesty over $\mathcal{M}$ requires expected length of order r_n, in line with modulus-based accounts of optimal confidence sets [Low, 1997, Armstrong and Kolesár, 2016].

The two theorems above leave the rate in the compact form r_n. The next result expands that expression into the threshold regimes induced by the information-balance bandwidth. Two deterministic scales organize those regimes, the risk-transition scale $\delta_{\mathrm{crit},n}$ and the bandwidth-transition scale $\delta_{\mathrm{edge},n}$,

$$\delta_{\mathrm{crit},n} = n^{-1/\{2(\beta\kappa+2\beta+\kappa+1)\}}, \qquad \delta_{\mathrm{edge},n} = n^{-1/(2\beta+\kappa+1)},$$

and they are restated inside Theorem 3. Table 1 collects the regimes, separating the geometry of the treatment support at the threshold from the resulting risk order.

Theorem 3. [`thm:phase-diagram`] (Phase boundary regimes) *Let J be a positive integer and let*

$$\beta > 0, \qquad \kappa \geq 0, \qquad L > 0, \qquad 0 < c_- \leq \kappa+1 \leq c_+, \qquad 0 < p_{\min} \leq 1/J, \qquad 0 < \bar{\delta} < 1, \qquad 0 < \alpha < 1/2.$$

Define the phase and edge scales by

$$\delta_{\mathrm{crit},n} = n^{-1/\{2(\beta\kappa+2\beta+\kappa+1)\}}, \qquad \delta_{\mathrm{edge},n} = n^{-1/(2\beta+\kappa+1)}.$$

Then

$$\frac{\delta_{\mathrm{crit},n}}{\delta_{\mathrm{edge},n}} \to \infty.$$

Moreover, for every deterministic threshold sequence $(\delta_n)_{n\geq 1}$ with $\delta_n \in [0,\bar{\delta}]$ for every n, let h_n be the information-balance bandwidth in Definition 8, and set

$$a_n = \delta_n^{\kappa+1} h_n^{\beta}, \qquad r_n = n^{-1/2} + \delta_n^{\kappa+1} h_n^{\beta}$$

as in Definition 9. The following conclusions hold:

- ***Regular thresholds.*** *If $\delta_n/\delta_{\mathrm{crit},n} \to 0$, then $r_n \asymp n^{-1/2}$.*
- ***Critical thresholds.*** *For every $c_0 > 0$, if $\delta_n/\delta_{\mathrm{crit},n} \to c_0$, then $a_n \asymp n^{-1/2}$ and $r_n \asymp n^{-1/2}$.*

Result type	Condition	Treatment-support geometry	Conclusion
Bandwidth	$\delta_n = O(\delta_{\text{edge},n})$	window reaches the thinned endpoint $a = 0$	$h_n \asymp n^{-1/(2\beta+\kappa+1)}$.
Bandwidth	$\delta_n/\delta_{\text{edge},n} \to \infty$	window sits away from the thinned endpoint	$h_n \asymp (n\delta_n^\kappa)^{-1/(2\beta+1)}$.
Risk	$\delta_n/\delta_{\text{crit},n} \to 0$	threshold near the thinned endpoint	retained mean dominates: $r_n \asymp n^{-1/2}$.
Risk	$\delta_n/\delta_{\text{crit},n} \to c_0 \in (0,\infty)$	threshold near the thinned endpoint	retained mean and atom term balance: $r_n \asymp n^{-1/2}$.
Risk	$\delta_n/\delta_{\text{crit},n} \to \infty$ and $\delta_n \to 0$	threshold vanishing but above the edge scale	atom term dominates: $r_n \asymp \delta_n^{\kappa+1}(n\delta_n^\kappa)^{-\beta/(2\beta+1)}$.
Risk	$\delta_n \to \delta_0 \in (0,\bar\delta]$	threshold interior to the treatment support	atom-weighted pointwise Hölder rate: $r_n \asymp n^{-\beta/(2\beta+1)}$.
Risk	$\delta_n = 0$	no mass collapsed	$r_n = n^{-1/2}$.

Table 1: Bandwidth and risk regimes for the lower-threshold clamp, with $\delta_{\text{crit},n} = n^{-1/\{2(\beta\kappa+2\beta+\kappa+1)\}}$ and $\delta_{\text{edge},n} = n^{-1/(2\beta+\kappa+1)}$. The third column records the geometry of the treatment support at the threshold; the fourth records which contribution determines the risk order.

- ***Vanishing atom-dominated thresholds.*** *If $\delta_n/\delta_{\text{crit},n} \to \infty$ and $\delta_n \to 0$, then*

$$r_n \asymp \delta_n^{\kappa+1}\left(n\delta_n^\kappa\right)^{-\beta/(2\beta+1)}.$$

- ***Fixed thresholds.*** *For every $\delta_0 \in (0,\bar\delta]$, if $\delta_n \to \delta_0$, then*

$$r_n \asymp n^{-\beta/(2\beta+1)}.$$

- ***Zero threshold identities.*** *For every n and every $P \in \mathcal{M}$ satisfying Definition 5,*

$$q_x(0) = 0 \quad \textit{for every } x, \qquad \theta_0(P) = \int Y\,dP, \qquad r_n\big|_{\delta_n=0} = n^{-1/2},$$

where the last identity uses the bandwidth h_n at threshold 0.

- ***Zero threshold estimator.*** *For every n, every $P \in \mathcal{M}$ satisfying Definition 5, and every admissible three-way split $B = (I_0, I_1, I_2)$ as in Algorithm 1, the total-Gram estimator in Algorithm 1, formed at threshold 0 with bandwidth h_n, agrees under the product sampling law generated by P with*

$$z \mapsto \operatorname{clamp}_{[0,1]}\left(\frac{1}{|I_0|}\sum_{i\in I_0} Y_i\right).$$

- ***Zero threshold stabilized guarantees.*** *For every deterministic sequence $B_\bullet = (B_n)_{n\geq 1}$ of admissible three-way split blocks, the stabilized zero-threshold procedure has coverage at* *least $1-\alpha$ for all sufficiently large n. Its stabilized worst-case risk at threshold 0 and its stabilized worst-case expected length at threshold 0 with noncoverage level α are both asymptotic to $n^{-1/2}$.*

Theorem 3 identifies $\delta_{\mathrm{crit},n}$ as the boundary at which the local atom contribution reaches the parametric scale, and $\delta_{\mathrm{edge},n}$ as the smaller edge scale for the bandwidth algebra. Regular and critical threshold paths retain root-n order. Vanishing thresholds above the critical scale inherit a degenerate-design nonparametric rate of the kind studied by Gaïffas [2005]: the estimation point rides toward the thinned endpoint $a=0$, so the local design mass degenerates at the polynomial rate implied by Assumption 4. A fixed positive threshold is an interior-design geometry. There the target point δ_0 is interior to the treatment support and the density at that point is bounded away from zero, so the exponent $\beta/(2\beta+1)$ is the ordinary positive-density pointwise Hölder rate [Stone, 1982, Donoho and Liu, 1991a] multiplied by an atom weight $\delta_0^{\kappa+1}$ bounded away from zero. The estimator still uses a one-sided window, because the clamp evaluates the regression at δ_0 from above, but that is a property of the procedure rather than of the design geometry. At the exact zero threshold, the clamp target reduces to the observed mean and the total-Gram estimator reduces to the clipped split-sample outcome mean.

The section closes with a scalar calibration. It specializes the localized perturbation to one stratum, linear thinning, and Lipschitz smoothness, giving an explicit exponent check for the critical scale.

Proposition 1. [`prop:one-cell-calibration`] (One-cell calibration) *Let $\bar{\delta}\in(0,1)$, and let $(\delta_n)_{n\ge 1}$ be a deterministic threshold sequence satisfying*

$$0\le \delta_n\le \bar{\delta}\qquad \textit{for every } n.$$

Assume that $\delta_n \gg n^{-1/4}$, in the sense that

$$\frac{\delta_n}{n^{-1/4}}\longrightarrow\infty.$$

In the one-stratum model with $\beta=\kappa=1$ and $\pi(a)=2a$, let h_n be the information-balance bandwidth from Definition 8 evaluated at $(\delta_n,1,1,\bar{\delta})$. For the concrete localized Bernoulli alternatives, define

$$K_n = n\int_0^1 \mathrm{KL}(\mathrm{Bernoulli}(m_{1,\delta_n,h_n}(a)),\mathrm{Bernoulli}(m_0(a)))\ 2a\,da$$

and

$$\Delta_n=\int_{\delta_n}^1\big(m_{1,\delta_n,h_n}(a)-m_0(a)\big)\,2a\,da+\left(\int_0^{\delta_n}2a\,da\right)\big(m_{1,\delta_n,h_n}(\delta_n)-m_0(\delta_n)\big).$$

Then

$$h_n\asymp(n\delta_n)^{-1/3},\qquad K_n\asymp n\delta_n h_n^3,\qquad \Delta_n\asymp\delta_n^2h_n.$$

Moreover,

$$\delta_n\asymp n^{-1/10}\quad\Longleftrightarrow\quad\Delta_n\asymp n^{-1/2}.$$

In Proposition 1, K_n is the product Kullback–Leibler scale for the localized Bernoulli alternative, while Δ_n is the induced clamp-target separation. The balance $h_n\asymp(n\delta_n)^{-1/3}$ keeps the localized alternative statistically contiguous at constant information scale, and the identity $\delta_n\asymp n^{-1/10}$ reproduces the critical exponent from Theorem 3 when $\beta=\kappa=1$.

6 Causal interpretation through full-data lifts

The observed-data results in Theorems 1 and 2 concern the clamp functional $\theta_\delta(P)$ under the law of (X, A, Y). This section supplies a full-data causal interpretation for the same target. One notational convention applies throughout this section and the continuity-only section. Every estimator and interval used in this paper is a measurable function of the observed sample, so its risk, coverage probability, and expected length are the same whether computed under the n-fold product law of the observed margin or under the n-fold product law of the full-data package. We therefore state all such quantities under the observed-margin product law $P^{\otimes n}$, and use the full-data product law only where the statement genuinely concerns full-data objects. The structural full-data overlay augments the observed experiment with a full-data law P^{F}, a member-specific latent-response space $\mathcal{U}$, a latent response variable U, structural response maps g_x, and a potential-outcome process $\{Y(a) : a \in [0,1]\}$. Within that overlay, the lower-threshold policy changes the realized treatment through the same clamp map introduced in Definition 6; the statistical analysis remains based on the observed sample.

The analysis rests on three full-data restrictions and one response-mean definition. The first two restrictions are standard causal conditions for a latent-response representation, and the third records the continuity required on the threshold interval used by the clamp policy.

Assumption 6. [`ass:consistency`] (Structural consistency) *Under P^{F}, the following conditions hold:*

- ***(Latent response.)*** *The latent response variable U takes values in $\mathcal{U}$.*
- ***(Structural maps.)*** *For every $x \in \mathcal{X}$, the map*

$$g_x : [0,1] \times \mathcal{U} \to [0,1]$$

  is jointly Borel measurable.
- ***(Potential and observed outcomes.)*** *There is a single P^{F}-null set outside which, for every $a \in [0,1]$,*

$$Y(a) = g_X(a, U), \qquad Y = g_X(A, U).$$

Assumption 6 is the structural consistency condition: a common latent response generates all potential outcomes and the observed outcome at the realized treatment, in the simultaneous NPSEM-style formulation used for modified treatment policies [Díaz et al., 2021]. The common null set makes the equality simultaneous over doses, which is the form needed when the clamp map evaluates the response at $d_\delta(A)$.

Assumption 7. [`ass:exchangeability`] (Conditional exchangeability) *Under P^{F}, the treatment A and latent response variable U are conditionally independent given X.*

Assumption 7 is the latent-response randomization condition [Díaz et al., 2021]. Within each stratum, it lets the conditional treatment density weight the same latent-response distribution that determines the structural mean.

Definition 19. [`synth_1`] (Full-data response mean) *For a full-data law package P^{F} with latent-response carrier $\mathcal{U}$, latent variable U, structural maps $g_x : [0,1] \times \mathcal{U} \to [0,1]$, and observed stratum X, the full-data response mean in stratum $x \in \mathcal{X}$ is*

$$m_x^{\mathrm{F}}(a) = \mathbb{E}_{P^{\mathrm{F}}}\{g_x(a, U) \mid X = x\}, \qquad a \in [0,1].$$

When the law package is in the full-data continuity model, the map $a \mapsto m_x^{\mathrm{F}}(a)$ is continuous on $[0, \bar{\delta}]$ for every $x \in \mathcal{X}$.

The full-data response mean $m_x^{\mathrm{F}}(a)$ is the fiberwise structural analogue of the observed regression extension. It isolates the average structural response at dose a among units in stratum x, which is the quantity the bridge result compares with μ_x^P on the threshold interval.

Assumption 8. [`ass:full-data-response-continuity`] (Full-data response continuity) *Under the full-data law P^{F}, for every $x \in \mathcal{X}$, the full-data response mean $a \mapsto m_x^{\mathrm{F}}(a)$ is continuous on $[0, \bar{\delta}]$.*

Assumption 8 is specific to this analysis. It places the continuity requirement directly on the structural response mean over the range where clamping can create a threshold evaluation, matching the interval on which the observed regression extension is used.

These conditions define the full-data causal class and the decision criteria used below.

Definition 20. [`def:full-data-model-class`] (Full-data causal class $\mathcal{M}^{\mathrm{F}}$) *$\mathcal{M}^{\mathrm{F}}$ is the collection of full-data law packages P^{F} satisfying the following memberwise structure relative to Definition 5 and Assumptions 6 to 8.*

- ***(Latent response space.)*** *Each member carries its own standard Borel latent-response space $\mathcal{U}$ and an $\mathcal{U}$-valued random variable U.*
- ***(Structural response maps.)*** *Each member carries jointly Borel maps*

$$g_x : [0,1] \times \mathcal{U} \to [0,1], \qquad x \in \mathcal{X},$$

  and a jointly measurable potential-outcome process

$$\{Y(a) : a \in [0,1]\}.$$

- ***(Full-data law.)*** *Each member carries a probability law for*

$$(X, A, Y, U, \{Y(a) : a \in [0,1]\}).$$

- ***(Observed margin.)*** *The (X, A, Y)-margin is some $P \in \mathcal{M}$.*
- ***(Causal member conditions.)*** *Within the member's own carrier, the simultaneous latent-response consistency, latent randomization, and fiberwise structural-mean continuity on $[0, \bar{\delta}]$ conditions hold.*

The class is formed member by member, with membership determined on each member's own standard Borel carrier.

The full-data class $\mathcal{M}^{\mathrm{F}}$ keeps the observed law inside $\mathcal{M}$ while allowing each member to carry its own standard Borel latent-response space. This memberwise formulation is the natural setting for comparing causal and observed decision problems without imposing a common latent carrier across all laws.

Definition 21. [`def:causal-clamp-mean`] (Causal clamp mean $\psi_\delta(P^{\mathrm{F}})$) *For a full-data law P^{F} as in Definition 20, the clamp policy d_δ from Definition 6, and $0 \le \delta \le \bar{\delta}$, define the causal clamp mean by*

$$\psi_\delta(P^{\mathrm{F}}) = \mathbb{E}_{P^{\mathrm{F}}}[Y(d_\delta(A))] \,.$$

The causal clamp mean $\psi_\delta(P^{\mathrm{F}})$ is the mean outcome under the same lower-threshold modified treatment policy used to define the observed clamp functional in Definition 7. Thus the causal estimand changes the interpretation of the response while preserving the policy rule and threshold range.

Definition 22. [`def:causal-frontier-criteria`] (Causal frontier criteria $R^\star_{n,\mathrm{F}}$ and $L^\star_{n,\mathrm{F}}$) *For the causal clamp mean in Definition 21, define the causal minimax absolute-error risk over estimators measurable with respect to the observed sample $O_1,\dots,O_n$ by*

$$R^\star_{n,\mathrm{F}} = \inf_{\widehat{\psi}} \sup_{P^{\mathrm{F}}\in\mathcal{M}^{\mathrm{F}}} \mathbb{E}_{(P^{\mathrm{F}})^{\otimes n}} \left|\widehat{\psi} - \psi_{\delta_n}(P^{\mathrm{F}})\right|.$$

For intervals C_n measurable with respect to the same observed sample, define the causal minimax honest expected length by

$$L^\star_{n,\mathrm{F}} = \inf_{C_n:\, \inf_{P^{\mathrm{F}}\in\mathcal{M}^{\mathrm{F}}} (P^{\mathrm{F}})^{\otimes n}\{\psi_{\delta_n}(P^{\mathrm{F}})\in C_n\}\ge 1-\alpha} \ \sup_{P^{\mathrm{F}}\in\mathcal{M}^{\mathrm{F}}} \mathbb{E}_{(P^{\mathrm{F}})^{\otimes n}}\left[\mathrm{len}(C_n)\right].$$

The causal minimax risk $R^\star_{n,\mathrm{F}}$ and causal honest-length criterion $L^\star_{n,\mathrm{F}}$ compare observed-sample estimators and intervals under the full-data class through their induced observed-sample behavior. The procedures therefore use the same data as in the observed experiment while the target is the causal clamp mean.

The next proposition gives the identification step. It connects the structural mean under the full-data law to the observed regression extension and then identifies the causal clamp mean with the observed clamp target.

Proposition 2. [`prop:causal-bridge`] (Causal bridge identity) *Let $P^{\mathrm{F}}\in\mathcal{M}^{\mathrm{F}}$ satisfy Definition 20, let P be its observed (X,A,Y)-margin, and let $d_\delta(a)=\max\{a,\delta\}$ be the clamp map in Definition 6. Suppose the frontier constants satisfy*

- ***(Strata.)*** *$J\ge 1$.*
- ***(Smoothness and thinning.)*** *$\beta>0$, $\kappa\ge 0$, and $L>0$.*
- ***(Density and mass constants.)*** *$c_->0$, $c_-\le\kappa+1\le c_+$, $p_{\min}>0$, and $p_{\min}\le 1/J$.*
- ***(Threshold and coverage levels.)*** *$0<\bar{\delta}<1$ and $0<\alpha<1/2$.*

For $x\in\mathcal{X}$ and $a\in[0,1]$, define the full-data response mean

$$m_x^{\mathrm{F}}(a) = P^{\mathrm{F}}(X=x)^{-1}\int_{\{X=x\}} g_x(a,U)\,dP^{\mathrm{F}}.$$

Then, for every $x\in\mathcal{X}$ and every Borel set $B\subseteq[0,1]$,

$$P(X=x)^{-1}\int_{\{X=x,\ A\in B\}} Y\,dP^{\mathrm{F}} = \int_B m_x^{\mathrm{F}}(a)\pi_x(a)\,da.$$

Moreover, for every $x\in\mathcal{X}$ and every $a\in[0,\bar{\delta}]$,

$$m_x^{\mathrm{F}}(a)=\mu_x^P(a).$$

Consequently, for every $\delta \in [0, \bar{\delta}]$*, the following identities hold:*

$$d_\delta(a) \in \operatorname{supp}\{\pi_x(a)\, da \text{ on } [0,1]\} \quad \text{for each } x \in \mathcal{X} \text{ and for conditional-treatment-a.e. } a,$$

$$Y(d_\delta(A)) = Y\mathbf{1}\{\delta < A\} + Y(\delta)\mathbf{1}\{A \le \delta\} \quad P^{\mathrm{F}}\text{-a.s.},$$

and, with

$$\psi_\delta(P^{\mathrm{F}}) = \int Y(d_\delta(A))\, dP^{\mathrm{F}}, \qquad \nu_\delta(P) = \int Y\mathbf{1}\{\delta < A\}\, dP, \qquad q_x(\delta) = \int_0^\delta \pi_x(a)\, da,$$

the causal clamp mean equals the observed clamp target

$$\psi_\delta(P^{\mathrm{F}}) = \nu_\delta(P) + \sum_{x \in \mathcal{X}} p_x q_x(\delta) \mu_x^P(\delta) = \theta_\delta(P).$$

Proposition 2 is the fiberwise bridge. Conditional exchangeability converts observed outcomes within a stratum and treatment set into an integral of the structural response mean against the treatment density; continuity and the observed regression extension then identify that mean on the threshold interval. The final identity shows that the causal lower-threshold policy mean is exactly the clamp functional analyzed in the observed experiment.

A second structural fact records that the full-data class covers the observed model class exactly at the margin. The construction uses regular conditional distributions on standard Borel spaces, a standard measurable-disintegration tool [Kallenberg, 2002, Theorem 6.3].

Proposition 3. [`prop:observed-margin-surjectivity`] (Observed margin surjectivity)[1] *Fix* $J \in \mathbb{N}$ *and constants* $\beta, \kappa, L, c_-, c_+, p_{\min}, \bar{\delta}, \alpha \in \mathbb{R}$*. Assume:*

- ***(Regime constants.)*** $0 < J$, $0 < \beta$, $0 \le \kappa$, $0 < L$, $0 < c_-$, $c_- \le \kappa + 1$, $\kappa + 1 \le c_+$, $0 < p_{\min}$, $p_{\min} \le 1/J$, $0 < \bar{\delta} < 1$, *and* $0 < \alpha < 1/2$.
- ***(Observed model.)*** $P \in \mathcal{M}$ *means that* P *satisfies Definition 5 with constants* $(\beta, \kappa, L, c_-, c_+, p_{\min})$.
- ***(Full-data model.)*** $P^{\mathrm{F}} \in \mathcal{M}^{\mathrm{F}}$ *means that* P^{F} *satisfies Definition 20 with constants* $(\beta, \kappa, L, c_-, c_+, p_{\min}, \bar{\delta})$.

Then, for every $P \in \mathcal{M}$*, there exists* $P^{\mathrm{F}} \in \mathcal{M}^{\mathrm{F}}$ *whose observed* (X, A, Y)*-margin is exactly* P*. Hence the observed-margin map from* $\mathcal{M}^{\mathrm{F}}$ *onto* $\mathcal{M}$ *is surjective.*

Moreover, for every $n \in \mathbb{N}$ *and every* $\delta \in [0, \bar{\delta}]$*, the causal frontier criteria in Definition 22 satisfy*

$$R^\star_{n,\mathrm{F}}(\delta) = R^\star_n(\delta) \qquad \text{and} \qquad L^\star_{n,\mathrm{F}}(\delta, \alpha) = L^\star_n(\delta, \alpha).$$

Here $R^\star_n(\delta)$ *is the infimum, over observed-measurable estimators with values in* $[0,1]$*, of the worst-case absolute-error risk over* $P \in \mathcal{M}$ *for the observed clamp target at threshold* δ*, and* $L^\star_n(\delta, \alpha)$ *is the infimum, over uniformly honest observed-measurable confidence procedures at level* $1 - \alpha$*, of the worst-case expected interval length over* $P \in \mathcal{M}$*.*

[1] **Formalization scope.** This result uses the published conclusion [Kallenberg, 2002] (Theorem 6.3 (conditional distribution)). The cited source proof is not formalized here: Lean verifies its use and all remaining steps. If that cited conclusion is false, the portions of this result that depend on it are not certified.

Proposition 3 establishes the decision-theoretic equivalence needed for causal transport. Every observed law in $\mathcal{M}$ has a full-data representative in $\mathcal{M}^{\mathrm{F}}$, and the risk and honest-length criteria agree at each fixed threshold. Thus lower bounds and attainable procedures for the observed clamp target can be read as statements about the causal clamp mean on the declared full-data class.

One technical convention remains for comparing expected lengths. The next definition fixes the order-preserving way real bounds are read inside the extended nonnegative reals.

Definition 23. [`synth_11`] (Real-to-extended embedding) *Under the notation and hypotheses of the objects introduced earlier in the paper, a real number $t \in \mathbb{R}$ is embedded into the extended nonnegative reals by*

$$t \longmapsto \max\{t, 0\} \in [0, \infty] \,.$$

The image is finite and equals the positive part of t, so real expected-length bounds are compared with extended-real expected lengths through this embedding.

The embedding in Definition 23 lets the real constants in the rate bounds be compared with honest expected lengths when those lengths are represented in the extended nonnegative-real order. This is the scale used in the final causal length statement.

The main causal conclusion now follows by combining the bridge identity, observed-margin surjectivity, and the observed-data minimax and honest-interval results.

Theorem 4. [`thm:causal-frontier-lift`] (Causal frontier lift) *Let $J \in \mathbb{N}$, and let $\beta, \kappa, L, c_-, c_+, p_{\min}, \bar{\delta}, \alpha \in \mathbb{R}$. Suppose that*

- ***(Regime constants.)*** *The constants satisfy*

$$0 < J, \qquad \beta > 0, \qquad \kappa \geq 0, \qquad L > 0, \qquad c_- > 0, \qquad c_- \leq \kappa + 1 \leq c_+,$$

$$p_{\min} > 0, \qquad p_{\min} \leq \frac{1}{J}, \qquad 0 < \bar{\delta} < 1, \qquad 0 < \alpha < \frac{1}{2}.$$

- ***(Threshold sequence.)*** *The deterministic thresholds $(\delta_n)_{n \in \mathbb{N}}$ satisfy*

$$0 \leq \delta_n \leq \bar{\delta}$$

  for every $n \in \mathbb{N}$.

- ***(Split blocks.)*** *For each $n \in \mathbb{N}$, $B_n = (I_{0,n}, I_{1,n}, I_{2,n})$ is a deterministic three-way split with pairwise disjoint blocks and*

$$\left\lfloor \frac{n}{4} \right\rfloor \leq |I_{0,n}|, \qquad \left\lfloor \frac{n}{4} \right\rfloor \leq |I_{1,n}|, \qquad \left\lfloor \frac{n}{4} \right\rfloor \leq |I_{2,n}|.$$

Let h_n, the information-balance bandwidth for δ_n, be as in Definition 8, and let

$$r_n = n^{-1/2} + \delta_n^{\kappa+1} h_n^{\beta}$$

be the frontier rate from Definition 9. Then there exist constants $c, C \in \mathbb{R}$, with $0 < c < C$, depending only on these regime constants, such that, for every threshold sequence and every deterministic split-block sequence satisfying the preceding conditions, all sufficiently large n satisfy the following conclusions.

The total-Gram estimator $\widehat{\theta}_n$ in Algorithm 1, formed with split B_n, order $\ell = \lceil\beta\rceil - 1$, bandwidth h_n, and threshold δ_n, is observed-sample measurable. The corresponding bias-aware interval CI_n in Algorithm 2, formed with the same split, order, bandwidth, and threshold, is an observed-sample measurable interval.

For the causal minimax criteria $R^\star_{n,\mathrm{F}}$ and $L^\star_{n,\mathrm{F}}$ in Definition 22,

$$cr_n \le R^\star_{n,\mathrm{F}} \le \sup_{P^{\mathrm{F}}\in\mathcal{M}^{\mathrm{F}}} \mathbb{E}_{(P^{\mathrm{F}})^{\otimes n}} \left|\widehat{\theta}_n - \psi_{\delta_n}(P^{\mathrm{F}})\right| \le Cr_n,$$

$$\inf_{P^{\mathrm{F}}\in\mathcal{M}^{\mathrm{F}}} (P^{\mathrm{F}})^{\otimes n}\{\psi_{\delta_n}(P^{\mathrm{F}}) \in \mathrm{CI}_n\} \ge 1-\alpha,$$

$$\sup_{P^{\mathrm{F}}\in\mathcal{M}^{\mathrm{F}}} \mathbb{E}_{(P^{\mathrm{F}})^{\otimes n}} \operatorname{len}(\mathrm{CI}_n) \le Cr_n.$$

Moreover, in the extended nonnegative-real order, after embedding the nonnegative real bounds into $\mathbb{R}_{\ge 0}\cup\{\infty\}$,

$$cr_n \le L^\star_{n,\mathrm{F}} \le Cr_n.$$

Theorem 4 transports the observed minimax and honest-length frontier to the causal clamp mean on $\mathcal{M}^{\mathrm{F}}$. The same total-Gram estimator and the same bias-aware interval remain observed-sample procedures, and their guarantees hold uniformly over the full-data class because Proposition 2 aligns the target while Proposition 3 aligns the induced decision criteria. The rate is therefore the same r_n characterized in Definition 9 and Theorem 3: root-n and boundary-driven regimes carry over with the causal interpretation supplied by the full-data lift.

7 Continuity-only clamp inference

Sections 3 and 5 impose a Hölder modulus for the regression at the lower threshold. We now record the enlarged continuity-only experiment, which keeps the same density, stratum-mass, and lower-tail thinning restrictions while replacing the Hölder regression member with qualitative continuity.

Definition 24. [`def:continuity-model-class`] (Continuity-only model class $\mathcal{M}_{\mathrm{cont}}$) *$\mathcal{M}_{\mathrm{cont}}$ is the set of laws*

$$P \in \mathcal{P}(\mathcal{X}\times[0,1]\times[0,1])$$

satisfying Assumptions 2 to 4 and the following continuity condition: for every $x\in\mathcal{X}$, an almost-everywhere version of

$$a \mapsto \mathbb{E}_P[Y \mid A=a, X=x]$$

admits a continuous extension to $[0,\bar{\delta}]$. The continuity extension is required memberwise for each stratum.

The class in Definition 24 preserves the design restrictions that govern the amount of mass near the lower endpoint. Its regression condition supplies a selected point value at the clamp threshold through continuity, rather than through a quantitative Hölder modulus.

The corresponding target has the same retained-course plus threshold-atom decomposition as $\theta_\delta(P)$ in Definition 7; the atom value is read from the continuous extension in the continuity-only class.

Definition 25. [`def:continuity-clamp-functional`] (Continuity clamp target $\theta_{\delta,\mathrm{cont}}(P)$) *For $P \in \mathcal{M}_{\mathrm{cont}}$ and $0 \le \delta \le \bar{\delta}$, with $q_x(\delta)$ as in Definition 7 and $\mathcal{M}_{\mathrm{cont}}$ as in Definition 24, define*

$$\theta_{\delta,\mathrm{cont}}(P) = \mathbb{E}_P[Y\mathbf{1}\{A > \delta\}] + \sum_{x\in\mathcal{X}} p_x q_x(\delta)\mu^P_{x,\mathrm{cont}}(\delta).$$

The continuity-only scale is determined by the ordinary sampling fluctuation and the amount of mass moved by the lower-threshold clamp.

Definition 26. [`def:continuity-frontier`] (Continuity frontier s_n) *Define the continuity-only frontier rate by*

$$s_n = n^{-1/2} + \delta_n^{\kappa+1}.$$

The estimator and interval make the enlarged model explicit. They use the retained-course mean above the threshold and a fixed midpoint value for the lower-tail atom, with a bounded-sum allowance for the empirical lower-tail mass.

Algorithm 3 (Continuity fallback estimator $\widehat{\theta}_{n,\mathrm{cont}}$)

The inputs are the deterministic split blocks I_0, I_1, the threshold δ_n, and the observations $O_1, \dots, O_n$.

1. Define the retained-course empirical mean

$$\widehat{\nu}_n = |I_0|^{-1} \sum_{i\in I_0} Y_i \mathbf{1}\{A_i > \delta_n\}.$$

2. For each $x \in \mathcal{X}$, define the empirical lower-threshold atom mass

$$\widehat{b}_x = |I_1|^{-1} \sum_{i\in I_1} \mathbf{1}\{X_i = x, A_i \le \delta_n\}.$$

3. Output the fixed-regression-fallback estimator

$$\widehat{\theta}_{n,\mathrm{cont}} = \Pi_{[0,1]}\left\{\widehat{\nu}_n + \frac{1}{2}\sum_{x\in\mathcal{X}} \widehat{b}_x\right\}.$$

Algorithm 4 (Continuity honest interval $\mathrm{CI}_{n,\mathrm{cont}}$)

The inputs are the split blocks I_0, I_1, the level α, the estimator $\widehat{\theta}_{n,\mathrm{cont}}$, and the empirical masses $\widehat{b}_x$.

1. For $j = 0, 1$, define the split-specific Hoeffding radius

$$t_{j,\alpha} = \sqrt{\frac{\log(4/\alpha)}{2|I_j|}}.$$

2. Define the empirical total lower-threshold atom mass

$$\widehat{B}_n = \sum_{x \in \mathcal{X}} \widehat{b}_x.$$

3. Output the Hoeffding interval

$$\mathrm{CI}_{n,\mathrm{cont}} = \left[\widehat{\theta}_{n,\mathrm{cont}} - t_{0,\alpha} - t_{1,\alpha} - \frac{1}{2}\widehat{B}_n,\ \widehat{\theta}_{n,\mathrm{cont}} + t_{0,\alpha} + t_{1,\alpha} + \frac{1}{2}\widehat{B}_n\right] \cap [0,1].$$

The concentration step for Algorithm 4 uses the standard bounded-sum inequality of Hoeffding [1963, Theorem 2], applied separately to the retained-course and atom-mass split statistics.

The causal version of the continuity-only analysis uses the same structural ingredients as Definition 20, with the observed margin placed in $\mathcal{M}_{\mathrm{cont}}$.

Definition 27. [`def:continuity-full-data-model-class`] (Continuity full-data class $\mathcal{M}^{\mathrm{F}}_{\mathrm{cont}}$) *Define $\mathcal{M}^{\mathrm{F}}_{\mathrm{cont}}$ as the collection of full-data law packages P^{F} with a memberwise standard Borel latent-response space $\mathcal{U}$, latent variable U, jointly Borel stratum-specific response maps g_x, and jointly measurable potential-outcome process $\{Y(a) : a \in [0,1]\}$, such that:*

- ***(Observed margin.)*** *The observed (X, A, Y)-margin of P^{F} belongs to $\mathcal{M}_{\mathrm{cont}}$ in Definition 24.*
- ***(Consistency.)*** *The package satisfies Assumption 6.*
- ***(Exchangeability.)*** *The package satisfies Assumption 7.*
- ***(Response continuity.)*** *The package satisfies Assumption 8.*

The carrier spaces $\mathcal{U}$ may vary with P^{F}.

For later comparison, the next definition fixes the observed margin of a full-data law as the pushforward onto the observed coordinates.

Definition 28. [`def:continuity-frontier-criteria`] (Continuity frontier criteria $R^\star_{n,\mathrm{cont}}$ and $L^\star_{n,\mathrm{cont}}$) *For estimators and intervals measurable with respect to $O_1, \ldots, O_n$, define the continuity-only observed minimax absolute-error risk by*

$$R^\star_{n,\mathrm{cont}} = \inf_{\widehat{\theta}} \sup_{P \in \mathcal{M}_{\mathrm{cont}}} \mathbb{E}_{P^{\otimes n}} \left|\widehat{\theta} - \theta_{\delta_n,\mathrm{cont}}(P)\right|.$$

Define the continuity-only observed minimax honest expected length by

$$L^\star_{n,\mathrm{cont}} = \inf_{C_n : \inf_{P \in \mathcal{M}_{\mathrm{cont}}} P^{\otimes n}\{\theta_{\delta_n,\mathrm{cont}}(P) \in C_n\} \geq 1-\alpha} \ \sup_{P \in \mathcal{M}_{\mathrm{cont}}} \mathbb{E}_{P^{\otimes n}}[\mathrm{len}(C_n)].$$

For the full-data continuity-only class, define the causal minimax absolute-error risk by

$$R^\star_{n,\mathrm{cont,F}} = \inf_{\widehat{\psi}} \sup_{P^{\mathrm{F}} \in \mathcal{M}^{\mathrm{F}}_{\mathrm{cont}}} \mathbb{E}_{(P^{\mathrm{F}})^{\otimes n}} \left|\widehat{\psi} - \psi_{\delta_n}(P^{\mathrm{F}})\right|.$$

Define the continuity-only causal minimax honest expected length by

$$L^\star_{n,\mathrm{cont,F}} = \inf_{C_n : \inf_{P^{\mathrm{F}} \in \mathcal{M}^{\mathrm{F}}_{\mathrm{cont}}} (P^{\mathrm{F}})^{\otimes n}\{\psi_{\delta_n}(P^{\mathrm{F}}) \in C_n\} \geq 1-\alpha} \ \sup_{P^{\mathrm{F}} \in \mathcal{M}^{\mathrm{F}}_{\mathrm{cont}}} \mathbb{E}_{(P^{\mathrm{F}})^{\otimes n}}[\mathrm{len}(C_n)].$$

The decision-theoretic criteria are recorded for both the observed continuity-only class and its full-data counterpart.

Proposition 4. [`prop:continuity-causal-bridge`] (Continuity causal bridge) *Let $J \in \mathbb{N}$, let $\kappa, c_-, c_+, p_{\min}, \bar{\delta} \in \mathbb{R}$, and let P^{F} be a full-data law package with observed margin P. Suppose that*

- ***(Full-data continuity model.)*** *$P^{\mathrm{F}} \in \mathcal{M}^{\mathrm{F}}_{\mathrm{cont}}$ with constants $(\kappa, c_-, c_+, p_{\min}, \bar{\delta})$, as in Definition 27.*
- ***(Design constants.)*** *$J > 0$, $\kappa \geq 0$, $c_- > 0$, $c_- \leq \kappa + 1 \leq c_+$, $0 < p_{\min} \leq 1/J$, and $0 < \bar{\delta} < 1$.*

For each $x \in \mathcal{X}$ and $a \in \mathbb{R}$, define the full-data response mean by

$$m^{\mathrm{F}}_x(a) = P^{\mathrm{F}}(X = x)^{-1} \int_{\{X=x\}} g_x(a, U)\, dP^{\mathrm{F}}.$$

Then, for every $x \in \mathcal{X}$ and every $a \in [0, \bar{\delta}]$,

$$m^{\mathrm{F}}_x(a) = \mu^P_{x,\mathrm{cont}}(a).$$

Moreover, if

$$\psi_\delta(P^{\mathrm{F}}) = \int Y(d_\delta(A))\; dP^{\mathrm{F}},$$

then, for every $\delta \in [0, \bar{\delta}]$,

$$\psi_\delta(P^{\mathrm{F}}) = \theta_{\delta,\mathrm{cont}}(P),$$

where $\theta_{\delta,\mathrm{cont}}(P)$ is the continuity-only clamp functional in Definition 25.

The first causal statement aligns the full-data response mean with the observed continuous extension and therefore identifies the causal clamp mean with the continuity-only observed target.

Proposition 5. [`prop:continuity-observed-margin-surjectivity`] (Observed-margin surjectivity)[2] *Let $J \in \mathbb{N}$ and let $\kappa, c_-, c_+, p_{\min}, \bar{\delta}, \alpha \in \mathbb{R}$. Suppose that the continuity-regime design constants used in Definitions 24, 27 and 28 hold for $(J, \kappa, c_-, c_+, p_{\min}, \bar{\delta})$, and that $0 < \alpha < 1/2$.*

Then the observed-margin map from $\mathcal{M}^{\mathrm{F}}_{\mathrm{cont}}$ onto $\mathcal{M}_{\mathrm{cont}}$ is surjective: for every $P \in \mathcal{M}_{\mathrm{cont}}$, there exists $P^{\mathrm{F}} \in \mathcal{M}^{\mathrm{F}}_{\mathrm{cont}}$ whose observed margin is P. Moreover, for every $n \in \mathbb{N}$ and every $\delta \in [0, \bar{\delta}]$, the four continuity-frontier criteria of Definition 28, evaluated at $(J, n, \kappa, c_-, c_+, p_{\min}, \bar{\delta}, \delta, \alpha)$, satisfy

$$R^\star_{n,\mathrm{cont},\mathrm{F}} = R^\star_{n,\mathrm{cont}}, \qquad L^\star_{n,\mathrm{cont},\mathrm{F}} = L^\star_{n,\mathrm{cont}}.$$

Proposition 4 shows that qualitative continuity turns almost-everywhere regression information into the pointwise threshold value needed after the clamp collapses the lower tail. Under the stated full-data continuity conditions, the structural response mean agrees with the observed

[2]**Formalization scope.** This result uses the published conclusion [Kallenberg, 2002] (Theorem 6.3 (conditional distribution)). The cited source proof is not formalized here: Lean verifies its use and all remaining steps. If that cited conclusion is false, the portions of this result that depend on it are not certified.

continuous extension throughout the threshold envelope, and the causal clamp mean equals the observed continuity clamp target.

The next statement records that the continuity-only full-data class covers the observed continuity-only class exactly at the margin. Its construction uses regular conditional distributions on standard Borel spaces, as supplied by Kallenberg [2002, Theorem 6.3].

Definition 29. [synth_13] (Observed margin of a full-data law) *For a full-data law package P^{F} governing $(X, A, Y, U, \{Y(a) : a \in [0,1]\})$, define $P^{\mathrm{F}}_{\mathrm{obs}}$ to be the observed-data marginal law of (X, A, Y). Equivalently, if $\pi_{\mathrm{obs}}(X, A, Y, U, \{Y(a)\}) = (X, A, Y)$ is the coordinate projection onto $\mathcal{X} \times [0,1] \times [0,1]$, then $P^{\mathrm{F}}_{\mathrm{obs}} = (\pi_{\mathrm{obs}})_{\#} P^{\mathrm{F}}$.*

Proposition 5 puts the observed and causal continuity-only decision problems on the same footing. Every observed continuity-only law has a full-data representative satisfying the causal restrictions, and the minimax risk and honest expected-length criteria agree exactly for each sample size and threshold in the stated range.

The main continuity-only conclusion now combines the bridge identity, observed-margin surjectivity, and the fallback estimator and interval.

Theorem 5. [thm:continuity-only-frontier] (Continuity frontier characterization)[3] *Fix $J \in \mathbb{N}$ and constants $\kappa, c_-, c_+, p_{\min}, \bar{\delta}, \alpha \in \mathbb{R}$. Suppose*

$$0 < J, \qquad 0 \le \kappa, \qquad 0 < c_-, \qquad c_- \le \kappa{+}1 \le c_+, \qquad 0 < p_{\min} \le \frac{1}{J}, \qquad 0 < \bar{\delta} < 1, \qquad 0 < \alpha < \frac{1}{2}.$$

Then there exist constants $c, C \in \mathbb{R}$ with $0 < c < C$ such that the following statements hold.

- ***(Uniform frontier bounds.)*** *For every deterministic threshold sequence $(\delta_n)_{n\ge1}$ with $\delta_n \in [0, \bar{\delta}]$ for all n, and for every deterministic sequence $B_\bullet = (B_n)_{n\ge1}$ of admissible three-way split blocks, all sufficiently large n satisfy the following bounds. With*

  $$s_n = n^{-1/2} + \delta_n^{\kappa+1},$$

  and with $R^\star_{n,\mathrm{cont}}$, $L^\star_{n,\mathrm{cont}}$, $R^\star_{n,\mathrm{cont,F}}$, and $L^\star_{n,\mathrm{cont,F}}$ as in Definition 28,

  $$c s_n \le R^\star_{n,\mathrm{cont}} \le \sup_{P \in \mathcal{M}_{\mathrm{cont}}} \mathbb{E}_{P^{\otimes n}} \left| \widehat{\theta}_{n,\mathrm{cont}} - \theta_{\delta_n,\mathrm{cont}}(P) \right| \le C s_n,$$

  the interval $\mathrm{CI}_{n,\mathrm{cont}}$ has uniform coverage at least $1-\alpha$ over $\mathcal{M}_{\mathrm{cont}}$, its worst-case expected length is at most $C s_n$, and

  $$c s_n \le L^\star_{n,\mathrm{cont}}.$$

  Moreover,

  $$R^\star_{n,\mathrm{cont,F}} = R^\star_{n,\mathrm{cont}}, \qquad L^\star_{n,\mathrm{cont,F}} = L^\star_{n,\mathrm{cont}},$$

  and the same estimator and interval satisfy

  $$R^\star_{n,\mathrm{cont,F}} \le \sup_{P^{\mathrm{F}} \in \mathcal{M}^{\mathrm{F}}_{\mathrm{cont}}} \mathbb{E}_{(P^{\mathrm{F}}_{\mathrm{obs}})^{\otimes n}} \left| \widehat{\theta}_{n,\mathrm{cont}} - \psi_{\delta_n}(P^{\mathrm{F}}) \right| \le C s_n,$$

  with causal coverage at least $1-\alpha$ uniformly over $\mathcal{M}^{\mathrm{F}}_{\mathrm{cont}}$ and causal worst-case expected length at most $C s_n$.

- ***(Elbow and null-threshold rates.)*** *For the continuity-only rate in Definition 26,*

$$s_n\big(\delta_n = n^{-1/(2(\kappa+1))}\big) \asymp n^{-1/2}, \qquad s_n(0) \asymp n^{-1/2}.$$

- ***(Fixed positive thresholds.)*** *For every deterministic threshold sequence $(\delta_n)_{n\geq 1}$ with $\delta_n \in [0, \bar{\delta}]$ for all n, every $\delta_0 \in (0, \bar{\delta}]$, and $\delta_n \to \delta_0$,*

$$R^{\star}_{n,\mathrm{cont}} \not\to 0, \qquad L^{\star}_{n,\mathrm{cont}} \not\to 0.$$

Theorem 5 characterizes the continuity-only rate for both observed and causal formulations. The fallback estimator and Hoeffding interval attain the rate s_n, while the minimax risk and honest expected length are bounded below at the same order. The honest-length comparison follows the nonparametric confidence-interval lower-bound phenomenon of Low [1997], specialized here to the threshold-atom decision problem.

The rate has a simple interpretation. When δ_n is at the elbow $n^{-1/[2(\kappa+1)]}$, the lower-tail mass term $\delta_n^{\kappa+1}$ matches $n^{-1/2}$, so the continuity-only problem has a root-n scale. At $\delta_n = 0$, the atom term vanishes and the bounded retained-course mean again gives the root-n scale. For thresholds converging to a fixed positive value, the theorem records persistent minimax risk and honest length, reflecting the positive amount of treatment mass assigned a threshold regression value governed by qualitative continuity.

8 Discussion, limitations, and open questions

The results characterize how lower-tail overlap and smoothness jointly determine inference for threshold modified treatment policies with continuous treatments. In the Hölder model $\mathcal{M}$, Theorems 1 and 2 give the common rate r_n, combining the ordinary sampling term $n^{-1/2}$ with the local extrapolation term $\delta_n^{\kappa+1} h_n^{\beta}$. The phase diagram in Theorem 3 explains how this expression changes with the threshold sequence: moving the clamp closer to the lower endpoint changes both the mass affected by the policy and the amount of information available near the threshold.

The continuity-only comparison in Theorem 5 separates quantitative smoothness from qualitative continuity. With a Hölder radius and exponent fixed in the model, the local polynomial construction uses observations in a shrinking neighborhood of δ_n and pays the bias term h_n^{β}. With continuity alone, the delivered rate is $s_n = n^{-1/2} + \delta_n^{\kappa+1}$, because the threshold regression contribution is controlled through boundedness and continuity rather than through a quantitative modulus. Taken together, the two frontiers show that atom mass and smoothness enter different parts of the problem: the mass below the threshold weights the regression value to be learned, while the smoothness condition determines how sharply that value can be inferred from nearby continuous-treatment observations.

The causal lifting results give the same statistical rates a full-data interpretation. Under the consistency, exchangeability, and response-continuity conditions in Assumptions 6 to 8, Proposition 2 and Theorem 4 transfer the observed-data Hölder frontier to the causal clamp mean. The continuity-only bridge in Proposition 4 and Theorem 5 gives the analogous interpretation for the qualitative-continuity class. These statements make the role of the full-data class explicit: once

[3] **Formalization scope.** This result uses the published conclusion [Hoeffding, 1963] (Theorem 2, specialized to unit ranges and two tails). The cited source proof is not formalized here: Lean verifies its use and all remaining steps. If that cited conclusion is false, the portions of this result that depend on it are not certified.

the observed margin belongs to the relevant observed model and the stated causal restrictions hold, the observed clamp functional and the causal clamp mean have matching minimax criteria.

The exact-modulus perspective below records a more calibrated object behind the order results. The intervals in Algorithm 2 use analytic upper bounds on local bias and stochastic error. A sharper implementation can instead optimize directly over affine weights in the realized local design, in the spirit of modulus-based honest inference [Armstrong and Kolesár, 2018, Armstrong and Kolesár, 2016]. The next three definitions name the realized affine-modulus handle, the integrated split-specific radius, and the resulting interval used to formulate the sharp calibration question.

Definition 30. [`def:exact-modulus-handle`] (Exact modulus radius) *With notation as in Definitions 5 and 12, fix $P \in \mathcal{M}$ under the displayed $\beta, \kappa, L, c_-, c_+, p_{\min}$ member conditions, a deterministic split $B = (I_0, I_1, I_2)$ with pairwise-disjoint blocks satisfying $|I_m| \geq \lfloor n/4 \rfloor$ for $m = 0, 1, 2$, a stratum $x \in \mathcal{X}$, a local-polynomial degree $\ell \in \mathbb{N}$, a threshold δ, a bandwidth $h > 0$, and a multiplier $t \geq 0$. For a realized observed sample $z = (O_1, \ldots, O_n)$, with $O_i = (X_i, A_i, Y_i)$, set*

$$u_i = \frac{A_i - \delta}{h}, \qquad S_{z,x} = \{i \in I_2 : X_i = x,\ 0 \leq u_i \leq 1\}.$$

The affine weight set is

$$\mathcal{W}_{z,x} = \left\{ w = (w_i)_{i=1}^n \in \mathbb{R}^n : w_i = 0 \text{ whenever } i \notin I_2 \text{ or } X_i \neq x \text{ or } u_i \notin [0,1], \quad \sum_{i \in I_2} w_i u_i^j = \mathbf{1}\{j = 0\} \text{ for } j = 0, \ldots \right.$$

For $w \in \mathcal{W}_{z,x}$, let $B^{\mathrm{H}}_{z,x}(w)$ denote the exact Hölder bias term from the (β, L)-Hölder regression class in Assumption 5. Define the realized affine-modulus radius by

$$R^{\mathrm{ex}}_{z,x}(t) = \inf \left\{ B^{\mathrm{H}}_{z,x}(w) + t \left(\sum_{i \in I_2} w_i^2 \right)^{1/2} : w \in \mathcal{W}_{z,x} \right\},$$

with the real-valued convention that this infimum is 0 when the displayed feasible set is empty. The integrated exact-modulus radius is

$$M^{\mathrm{ex}}_x(P, B, \ell, \beta, \kappa, L, c_-, c_+, p_{\min}, \delta, h, t) = \int R^{\mathrm{ex}}_{z,x}(t)\, \mathrm{d}P^{\otimes n}(z),$$

where $P^{\otimes n}$ is the canonical product law of the observed sample. Both $R^{\mathrm{ex}}_{z,x}(t)$ and $M^{\mathrm{ex}}_x(P, B, \ell, \beta, \kappa, L, c_-, c_+, p_{\min}, \delta, h, t)$ are radius, or half-length, quantities.

Definition 31. [`synth_21`] (Exact-modulus radius W_{n,B_n}) *Adopt the notation and hypotheses of the objects introduced earlier in the paper. Fix $P \in \mathcal{M}$, a deterministic third split block I_2, a stratum $x \in \mathcal{X}$, and a multiplier $t \geq 0$. For a realized design $z = \{(X_i, A_i) : i \in I_2\}$, set*

$$u_i = \frac{A_i - \delta_n}{h_n}, \qquad S_{z,x} = \{i \in I_2 : X_i = x,\ 0 \leq u_i \leq 1\}.$$

Let

$$\mathcal{W}_{z,x} = \left\{ w = (w_i)_{i \in I_2} \in \mathbb{R}^{I_2} : w_i = 0 \text{ for } i \notin S_{z,x},\ \sum_{i \in I_2} w_i u_i^j = \mathbf{1}\{j = 0\} \text{ for } j = 0, \ldots, \ell \right\}.$$

Let $\mathcal{H}_{\beta,L}$ be the class of continuous functions $f : [0,1] \to [0,1]$ satisfying the declared order-ℓ Hölder Taylor-remainder bound with exponent β and radius L. For $w \in \mathcal{W}_{z,x}$, define the realized Hölder bias functional

$$B_{z,x}(w) = \sup_{f \in \mathcal{H}_{\beta,L}} \left| \sum_{i \in I_2} w_i f(A_i) - f(\delta_n) \right|.$$

The realized affine-modulus radius at multiplier t is

$$\mathcal{R}_{z,x}(t) = \begin{cases} \inf_{w \in \mathcal{W}_{z,x}} \left\{ B_{z,x}(w) + t \left(\sum_{i \in I_2} w_i^2 \right)^{1/2} \right\}, & \mathcal{W}_{z,x} \neq \varnothing, \\ 0, & \mathcal{W}_{z,x} = \varnothing. \end{cases}$$

The integrated exact-modulus radius is

$$\mathfrak{R}_{n,x}(P, I_2; t) = \int \mathcal{R}_{z,x}(t) \, \mathrm{d}(P_{X,A}^{\otimes I_2})(z),$$

where $P_{X,A}^{\otimes I_2}$ is the independent observed-design product law on the coordinates in I_2.

If $B_n = (I_{0,n}, I_{1,n}, I_{2,n})$ is an admissible deterministic three-way split at sample size n, its worst integrated exact-modulus radius is

$$W_{n,B_n} = \sup_{P \in \mathcal{M}} \sup_{x \in \mathcal{X}} \mathfrak{R}_{n,x}(P, I_{2,n}; t_\alpha), \qquad t_\alpha = \sqrt{\log(12J/\alpha)/2}.$$

Thus W_{n,B_n} is a radius quantity computed from the third block of B_n, and the associated untruncated symmetric full length is $2W_{n,B_n}$.

Definition 32. [`synth_20`] (Exact-modulus interval) *Assume the notation, model conditions, and preceding constructions in the objects introduced earlier in the paper. Fix an admissible deterministic three-way split $B_n = (I_0, I_1, I_2)$ at sample size n. For a realized observed sample z, let $\widehat{\theta}^{\mathrm{TG}}_{n,B_n}(z)$ be the total-Gram estimator formed with the blocks in B_n, and evaluate the realized affine-modulus radius $R^{\mathrm{ex}}_{z,x}(t_\alpha)$ on the (X, A)-projection of z using the third block I_2, where*

$$t_\alpha = \sqrt{\log(12J/\alpha)/2}.$$

Define

$$R_{n,B_n}(z) = \sup_{x \in \mathcal{X}} R^{\mathrm{ex}}_{z,x}(t_\alpha).$$

The exact-modulus interval is

$$C^{\mathrm{mod}}_{n,B_n}(z) = \left[\max\left\{0, \widehat{\theta}^{\mathrm{TG}}_{n,B_n}(z) - R_{n,B_n}(z)\right\}, \min\left\{1, \widehat{\theta}^{\mathrm{TG}}_{n,B_n}(z) + R_{n,B_n}(z)\right\} \right].$$

Its worst-case expected length over the observed-data model is

$$\ell^{\mathrm{mod}}_{n,B_n} = \sup_{P \in \mathcal{M}} \mathbb{E}_{P^{\otimes n}} \left[\mathrm{len}\left(C^{\mathrm{mod}}_{n,B_n} \right) \right].$$

Definitions 30 to 32 record one concrete, fully specified object against which the sharp-constant question can be posed: an interval centered at the total-Gram estimator whose half-width is the realized affine-modulus radius, intersected with $[0, 1]$. It is important to be precise about what that object does and does not calibrate. Its radius solves the affine-modulus problem for the threshold-regression component alone: it is the supremum over strata of unweighted realized regression radii, it does not include the retained-mean deviation radius, the atom-mass deviation radii, or the atom weights $p_x q_x(\delta_n)$ by which those components enter the target, and it does not aggregate across strata in the way the target does. On the singular realized designs where the affine constraints are infeasible the realized radius is zero, so the object is not by itself a coverage-guaranteed procedure; the honest interval with a proved coverage guarantee is Algorithm 2. The zero-threshold regime makes the gap explicit: there the target is an ordinary bounded mean with frontier $n^{-1/2}$, while this radius continues to solve a local-regression problem whose natural scale is h_n^{β}. The question below is therefore posed for a component-level modulus; extending it to a modulus for the complete clamp functional, with the regular components, atom weights, their uncertainty, stratum aggregation, and a singular-design fallback included, is the substantive part of the problem and is open. The integrated radius averages the design-dependent affine optimization under the observed-design product law for the third split block and is defined stratum by stratum, replacing the analytic threshold-regression radius by a realized optimization problem.

Limitations. The scope of the results is the fixed finite-stratum design described in Definition 5, with outcomes bounded in $[0, 1]$, known smoothness radius and exponent in the Hölder model, and a known polynomial-thinning exponent governing lower-tail overlap. The causal interpretation is attached to the declared full-data classes in Definitions 20 and 27, and the continuity-only conclusions are attached to the continuity-only observed and full-data classes in Definitions 24 and 27. Within that scope, the theorems characterize the minimax risk and honest expected-length orders for the lower-threshold clamp target and its causal lift. A further limitation concerns evidence rather than scope. The case made here for the stabilized estimator and the bias-aware interval is an asymptotic-order case, supported at the level of a single analytic calibration cell in Proposition 1. Practical questions about total-Gram stabilization on singular realized designs and interval behavior at moderate sample sizes concern quantities whose finite-sample behavior need not resemble the eventual regime: the frequency of the Gram fallback, the realized coverage, and the realized expected length all depend on constants this analysis does not track. A systematic finite-sample evaluation across the regular, critical, atom-dominated, fixed-threshold, and near-zero regimes, reporting those quantities and comparing against unstabilized local polynomial and regular modified-policy benchmarks where the targets and conditions permit a fair comparison, is left to future work and is not attempted here.

Open questions. Two questions remain open. The first is adaptation: the frontier established here is the known-class benchmark, with β, L, and κ fixed and available to the procedure, and whether the same frontier is attainable when those constants are unknown, and at what logarithmic cost, is not settled by the present analysis. The second concerns exact constants rather than rates. Modulus-based confidence-interval theory describes the difficulty of such problems through the local modulus of a functional over a smoothness class; Cai and Low [2004a] gives a general adaptation framework for confidence intervals, while Armstrong and Kolesár [2018], Armstrong and Kolesár [2016] develop honest inference procedures that use modulus calculations for bias-aware intervals. The honest expected-length lower bound proved here is obtained by direct two-point and localized-perturbation arguments, and it matches the risk frontier in order. The clamp target is linear in the outcome regression but includes the design-dependent weight $q_x(\delta_n)$, so sharp modulus calibration is treated as a separate question. The exact-modulus constructions in Definitions 30 to 32 identify one concrete candidate for studying sharp honest length: use the total-Gram center, compute the realized affine-modulus radius on the third split block, and compare its worst integrated radius to the minimax expected-length benchmark.

Remark 1. [`oeq:sharp-constant`] (Sharp length calibration) *The sharp constant question asks for the following property. For every*

$$J \in \mathbb{N}, \qquad \beta, \kappa, L, c_-, c_+, p_{\min}, \bar{\delta}, \alpha \in \mathbb{R}$$

satisfying

$$J > 0, \quad \beta > 0, \quad \kappa \geq 0, \quad L > 0, \quad c_- > 0, \quad c_- \leq \kappa{+}1 \leq c_+, \quad 0 < p_{\min} \leq \frac{1}{J}, \quad 0 < \bar{\delta} < 1, \quad 0 < \alpha < \frac{1}{2},$$

there is a constant $c > 0$ such that every deterministic threshold sequence $\delta_\bullet = (\delta_n)_{n \geq 1}$ with $0 \leq \delta_n \leq \bar{\delta}$ for all n, and whose phase-boundary ratio satisfies either

$$\frac{\delta_n}{\delta_{\mathrm{crit},n}} \longrightarrow q \quad \textit{for some } q \geq 0$$

or

$$\frac{\delta_n}{\delta_{\mathrm{crit},n}} \longrightarrow \infty,$$

has the following simultaneous calibration properties. First,

$$\frac{L_n^\star}{r_n} \longrightarrow c, \qquad r_n = n^{-1/2} + \delta_n^{\kappa+1} h_n^\beta,$$

where h_n is the information-balance bandwidth from Definition 8 and r_n is the frontier from Definition 9. Second, there exists a deterministic sequence $B_\bullet = (B_n)_{n \geq 1}$ of admissible three-way split blocks such that, for all sufficiently large n,

$$\inf_{P \in \mathcal{M}} P^{\otimes n} \left\{ \theta_{\delta_n}(P) \in C_{n,B_n}^{\mathrm{mod}} \right\} \geq 1 - \alpha,$$

where C_{n,B_n}^{mod} is the exact-modulus interval obtained from the split B_n using the exact-modulus construction of Definition 30. Its worst-case expected length satisfies

$$\frac{\ell_{n,B_n}^{\mathrm{mod}}}{r_n} \longrightarrow c.$$

Finally, the same sequence $B_\bullet$ admits least-favorable mixed-experiment witnesses: there are a constant $D > 0$, an amplitude $0 < \mathsf{amplitude} \leq 1/4$, and sequences

$$P_0, P_{\mathrm{reg}}, P_{\mathrm{loc}} : \mathbb{N} \to \mathcal{M}$$

such that, for all sufficiently large n, the three laws satisfy the clamp-model member conditions; their n-sample laws are the product laws in Assumption 1; the pair $(P_0(n), P_{\mathrm{reg}}(n))$ has the *declared global Bernoulli shift at scale $n^{-1/2}$; the pair $(P_0(n), P_{\mathrm{loc}}(n))$ has the declared localized Bernoulli perturbation at β, δ_n, h_n, and* amplitude*; and*

$$\chi^2\big(P_{\mathrm{reg}}(n)^{\otimes n}\big\|P_0(n)^{\otimes n}\big) \le D, \qquad \chi^2\big(P_{\mathrm{loc}}(n)^{\otimes n}\big\|P_0(n)^{\otimes n}\big) \le D.$$

With

$$\Delta_n^{\mathrm{mix}} = |\theta_{\delta_n}(P_{\mathrm{reg}}(n)) - \theta_{\delta_n}(P_0(n))| + |\theta_{\delta_n}(P_{\mathrm{loc}}(n)) - \theta_{\delta_n}(P_0(n))|,$$

and with W_{n,B_n} denoting the worst integrated exact-modulus radius from Definition 30 computed using the third block in B_n, the matching limits are

$$\frac{\Delta_n^{\mathrm{mix}}}{2W_{n,B_n}} \longrightarrow 1, \qquad \frac{\ell_{n,B_n}^{\mathrm{mod}}}{2W_{n,B_n}} \longrightarrow 1, \qquad \frac{2W_{n,B_n}}{L_n^\star} \longrightarrow 1.$$

A positive answer to Remark 1 would identify the sharp asymptotic full-length constant for the exact-modulus interval across the phase regimes covered by the displayed threshold condition. It would also connect the mixed regular-and-local perturbation comparison directly to the integrated realized radius, sharpening the order-optimal honest interval of Theorem 2 into an exact asymptotic calibration.

Appendices

A Proofs and verification note

This appendix collects the auxiliary statements that support the main risk, honest-length, causal-bridge, and continuity-only conclusions. The statements are organized by mathematical role: the clamp pushforward and support facts, the bandwidth and local-design algebra, the stabilized estimator and interval bounds, the phase calculations, the population moment identities underlying total-Gram stabilization, and the uniqueness statement for the continuity-only regression extension.

The first two propositions record the measure-theoretic behavior of the exact lower-threshold clamp. They connect the deterministic policy in Definition 6 to the mixed law that appears in the observed clamp functional in Definition 7.

Proposition 6. [`prop:pushforward-setup`] (Clamp pushforward decomposition) *Let J be a positive integer, let $P \in \mathcal{M}$ satisfy Definition 5, and let $d_\delta(a) = \max\{a, \delta\}$ as in Definition 6. Suppose the constants satisfy:*

- ***(Smoothness and thinning.)*** *$\beta > 0$, $\kappa \ge 0$, $L > 0$, and $0 < c_- \le \kappa + 1 \le c_+$.*
- ***(Stratum mass.)*** *$0 < p_{\min} \le 1/J$.*
- ***(Threshold and level.)*** *$0 < \bar\delta < 1$ and $0 < \alpha < 1/2$.*
- ***(Evaluation point.)*** *$x \in \mathcal{X}$ and $0 \le \delta \le \bar\delta$.*

Define the lower-tail atom mass

$$q_x(\delta) = \int_{[0,\delta]} \pi_x(a)\, da.$$

Then the conditional pushforward law of the clamped treatment satisfies

$$P\{d_\delta(A) \in da \mid X = x\} = P\{A \in da,\ A \in (\delta, 1] \mid X = x\} + q_x(\delta)\, \delta_\delta(da),$$

and the atom mass obeys

$$\frac{c_-\, \delta^{\kappa+1}}{\kappa+1} \le q_x(\delta) \le \frac{c_+\, \delta^{\kappa+1}}{\kappa+1}.$$

Proposition 6 supplies the exact mixed-measure decomposition induced by the clamp. The continuous part retains the natural treatment values above the threshold, and the lower tail becomes a point mass at the threshold whose order is $\delta^{\kappa+1}$ under Assumption 4. This atom-size calculation is the source of the weighting in Definition 9 and Definition 26.

Proposition 7. [`prop:policy-support`] (Policy support preservation) *Let P, $\beta, \kappa, L, c_-, c_+, p_{\min}, \bar{\delta}, \alpha$, x, and δ satisfy:*

- ***(Model.)*** *$P \in \mathcal{M}$ satisfies Definition 5 with constants $\beta, \kappa, L, c_-, c_+, p_{\min}$.*
- ***(Regime constants.)*** *$J \ge 1$, $0 < \beta$, $0 \le \kappa$, $0 < L$, $0 < c_- \le \kappa+1 \le c_+$, $0 < p_{\min} \le 1/J$, $0 < \bar{\delta} < 1$, and $0 < \alpha < 1/2$.*
- ***(Stratum and threshold.)*** *$x \in \mathcal{X}$ and $0 \le \delta \le \bar{\delta}$.*

Then, for conditional-treatment almost every a in stratum x, the clamped value $d_\delta(a) = \max\{a, \delta\}$ from Definition 6 is a support point of the conditional natural-treatment law given $X = x$: every open neighborhood U of $d_\delta(a)$ has positive conditional-treatment measure in stratum x,

$$0 < \Pr_P(A \in U \mid X = x).$$

Proposition 7 records the support property used by the causal bridge in Proposition 2. The polynomial lower envelope gives positive conditional treatment measure in every neighborhood of the clamp value, so the structural response at the clamped treatment is evaluated on the support generated by the observed design.

Lemmas 10 to 13 establish the bandwidth normalization and the realized local-polynomial algebra used by the total-Gram estimator. These statements translate the thinning envelope into local sample information and deterministic weight identities.

Lemma 1. [`lem:shifted-moment-coercivity`] (Uniform shifted coercivity) *Assume:*

- ***(Degree.)*** *Fix a polynomial degree $\ell \in \mathbb{N}$.*
- ***(Exponent.)*** *Fix a thinning exponent $\kappa \in \mathbb{R}$.*
- ***(Nonnegativity.)*** *The exponent satisfies $0 \le \kappa$, as in Assumption 4.*
- ***(Reference matrix.)*** *Let $M_{\ell,\kappa,\rho}$ denote the normalized population reference moment matrix in Definition 12.*

Then there exists a constant $c = c(\ell, \kappa) > 0$ such that, for every shift $\rho \ge 0$ and every vector $z = (z_0, \ldots, z_\ell) \in \mathbb{R}^{\ell+1}$,

$$c \sum_{j=0}^{\ell} z_j^2 \le z^\top M_{\ell,\kappa,\rho} z.$$

Proof of Lemma 1. For $z = (z_0, \dots, z_\ell)$, set

$$P_z(u) = \sum_{i=0}^{\ell} z_i u^i$$

and

$$E_\kappa(z) = \sum_{i=0}^{\ell} \sum_{j=0}^{\ell} \frac{z_i z_j}{i + j + \kappa + 1}.$$

Since $i + j + \kappa > -1$, expansion of the square and the identity $\int_0^1 u^{i+j+\kappa}\, du = (i + j + \kappa + 1)^{-1}$ give

$$E_\kappa(z) = \int_0^1 P_z(u)^2 u^\kappa \, du.$$

If $z \neq 0$, then P_z is a nonzero polynomial. Its zero set in $(0, 1)$ is finite, so choose $u_0 \in (0, 1)$ with $P_z(u_0) \neq 0$. At this point $P_z(u_0)^2 u_0^\kappa > 0$, while the integrand is continuous and nonnegative on $[0, 1]$. Hence

$$E_\kappa(z) > 0 \qquad (z \neq 0).$$

The map $z \mapsto E_\kappa(z)$ is continuous, because it is a finite sum of quadratic coordinate terms. Let

$$S = \{z \in \mathbb{R}^{\ell+1} : \|z\|_\infty = 1\}.$$

The set S is compact and nonempty, so E_κ attains its minimum on S, say at $z^\circ \in S$. Since $z^\circ \neq 0$, strict positivity of E_κ at nonzero vectors gives $E_\kappa(z^\circ) > 0$. Define

$$c = \frac{E_\kappa(z^\circ)}{\ell + 1}.$$

Then $c > 0$. For $z = 0$, the desired bound against $E_\kappa(z)$ is immediate. For $z \neq 0$, put

$$r = \|z\|_\infty, \qquad v = r^{-1} z.$$

Then $v \in S$, $E_\kappa(z) = r^2 E_\kappa(v)$, and

$$\sum_{j=0}^{\ell} z_j^2 \le (\ell + 1) r^2.$$

Therefore

$$c \sum_{j=0}^{\ell} z_j^2 \le E_\kappa(z^\circ) r^2 \le E_\kappa(v) r^2 = E_\kappa(z).$$

It remains to compare this unshifted energy with the normalized shifted moment matrix. Fix $\rho \ge 0$, and define

$$D_\rho = \int_0^1 (\rho + u)^\kappa \, du, \qquad N_\rho(z) = \int_0^1 P_z(u)^2 (\rho + u)^\kappa \, du.$$

By the definition of $M_{\ell,\kappa,\rho}$,

$$z^\top M_{\ell,\kappa,\rho} z = \frac{N_\rho(z)}{D_\rho}.$$

The denominator is positive. Indeed, $(\rho+u)^\kappa$ is continuous and nonnegative on $[0,1]$, and it is strictly positive near $u=1$. With

$$R_\rho = (\rho+1)^\kappa,$$

monotonicity of $t \mapsto t^\kappa$ on $[0,\infty)$ gives

$$D_\rho \le R_\rho.$$

For $u \in [0,1]$,

$$u(\rho+1) \le \rho + u,$$

and hence

$$R_\rho u^\kappa = \{u(\rho+1)\}^\kappa \le (\rho+u)^\kappa.$$

Multiplying by $P_z(u)^2$ and integrating yields

$$R_\rho E_\kappa(z) \le N_\rho(z).$$

Since $E_\kappa(z) \ge 0$, the inequalities $D_\rho \le R_\rho$ and $R_\rho E_\kappa(z) \le N_\rho(z)$ imply

$$E_\kappa(z) \le \frac{N_\rho(z)}{D_\rho}.$$

Combining the coercive bound for E_κ with the displayed expression for $z^\top M_{\ell,\kappa,\rho} z$ gives, for every $\rho \ge 0$,

$$c \sum_{j=0}^{\ell} z_j^2 \le z^\top M_{\ell,\kappa,\rho} z.$$

□

Lemma 10 turns the infimum definition of h_n into an eventual equality at the information boundary. This equality is the normalization used in the risk upper bound and in the expected weight-energy calculation.

Lemma 2. [`lem:lambda-star-positive`] (Positive Gram constant) *In the local-design notation of Definition 12, suppose:*

- ***(Degree.)*** *The polynomial degree is $\ell \in \mathbb{N}$.*
- ***(Thinning exponent.)*** *The exponent satisfies $\kappa \ge 0$, as in Assumption 4.*
- ***(Envelope constants.)*** *The envelope constants satisfy $c_- > 0$ and $c_+ > 0$.*

Then the uniform population Gram lower-bound constant

$$\lambda_\star(\ell,\kappa,c_-,c_+) = \frac{c_-}{c_+} \inf_{\rho \ge 0} \lambda_{\min}(M_{\ell,\kappa,\rho})$$

is strictly positive:

$$\lambda_\star(\ell,\kappa,c_-,c_+) > 0.$$

Proof of Lemma 2. By Lemma 1, there is a constant $c = c(\ell, \kappa) > 0$ such that, for every $\rho \geq 0$ and every $z \in \mathbb{R}^{\ell+1}$,

$$c \sum_{j=0}^{\ell} z_j^2 \leq z^\top M_{\ell,\kappa,\rho} z.$$

Taking the infimum over all unit vectors gives

$$c \leq \lambda_{\min}(M_{\ell,\kappa,\rho}) \qquad \text{for every } \rho \geq 0.$$

The indexing set $\{\rho : \rho \geq 0\}$ is nonempty, so

$$c \leq \inf_{\rho \geq 0} \lambda_{\min}(M_{\ell,\kappa,\rho}).$$

Because $c_- > 0$ and $c_+ > 0$,

$$\lambda_\star(\ell, \kappa, c_-, c_+) = \frac{c_-}{c_+} \inf_{\rho \geq 0} \lambda_{\min}(M_{\ell,\kappa,\rho}) \geq \frac{c_-}{c_+} c > 0.$$

□

Lemma 11 gives the population conditioning bound for the local monomial design. The lower bound is scaled by the local window mass, so it remains compatible with the global polynomial-thinning envelope in the lower-endpoint window.

Lemma 3. [`lem:power-window-integral-lower`] (Power window integral lower bound) *Using the thinning-exponent notation of Assumption 4, let δ be a threshold, h a window width, and κ an exponent. Assume:*

- ***(Threshold.)*** $0 \leq \delta$.
- ***(Window width.)*** $0 < h$.
- ***(Exponent.)*** $0 \leq \kappa$.

Then

$$\frac{h}{2} \left(\frac{\delta + h}{2} \right)^\kappa \leq \int_\delta^{\delta+h} a^\kappa \, da.$$

Proof of Lemma 3. Set

$$m = \delta + \frac{h}{2}.$$

Then

$$\delta \leq m \leq \delta + h, \qquad 0 \leq m, \qquad \frac{\delta + h}{2} \leq m.$$

For every $a \in [m, \delta + h]$, the hypotheses $0 \leq \delta$ and $0 < h$ give

$$0 \leq \frac{\delta + h}{2},$$

and the bound $(\delta + h)/2 \leq m$ recorded above, together with $m \leq a$, gives

$$\frac{\delta + h}{2} \leq m \leq a.$$

Together with $\kappa \geq 0$, monotonicity of real powers on nonnegative bases yields

$$\left(\frac{\delta+h}{2}\right)^{\kappa} \leq a^{\kappa}.$$

Thus

$$\int_{m}^{\delta+h} \left(\frac{\delta+h}{2}\right)^{\kappa} da \leq \int_{m}^{\delta+h} a^{\kappa}\, da.$$

The left-hand side is the constant integral over an interval of length $h/2$:

$$\int_{m}^{\delta+h} \left(\frac{\delta+h}{2}\right)^{\kappa} da = \frac{h}{2}\left(\frac{\delta+h}{2}\right)^{\kappa}.$$

Since $a^{\kappa} \geq 0$ on $[\delta, \delta+h]$ and $[m, \delta+h] \subseteq [\delta, \delta+h]$,

$$\int_{m}^{\delta+h} a^{\kappa}\, da \leq \int_{\delta}^{\delta+h} a^{\kappa}\, da.$$

Combining the window-length bound, the pointwise lower bound on a^{κ}, and the monotonicity of the window integral gives

$$\frac{h}{2}\left(\frac{\delta+h}{2}\right)^{\kappa} \leq \int_{\delta}^{\delta+h} a^{\kappa}\, da.$$

□

Lemma 12 is the deterministic local-polynomial fact used on the good-design event. Moment reproduction gives the intercept interpretation at the threshold, while the absolute-weight and squared-weight bounds control the Hölder bias and stochastic radius in Algorithm 2.

Lemma 4. [`lem:local-window-mass`] (Local window mass identity) *Let P, an observed-data law, belong to the model class $\mathcal{M}$ of Definition 5, with conditional treatment densities, stratum masses, and polynomial thinning as in Assumptions 2 to 4 for constants $(\beta, \kappa, L, c_-, c_+, p_{\min})$. Fix:*

- ***(Stratum.)*** *A stratum $x \in \mathcal{X}$.*
- ***(Threshold and width.)*** *A threshold $\delta \in \mathbb{R}$ and a window width $h \in \mathbb{R}$ with $h > 0$.*
- ***(Window support.)*** $[\delta, \delta+h] \subseteq [0, 1]$.

Then the population mass of the local window in stratum x is

$$\int \mathbf{1}\{X = x,\ 0 \leq (A-\delta)/h \leq 1\}\, dP = p_x \int_{\delta}^{\delta+h} \pi_x(a)\, da.$$

Proof of Lemma 4. Let

$$E_x(\delta, h) = \{O = (X, A, Y) : X = x,\ A \in [\delta, \delta+h]\}.$$

Because $h > 0$, for every observation O,

$$0 \leq \frac{A-\delta}{h} \leq 1 \quad \Longleftrightarrow \quad A \in [\delta, \delta+h].$$

Therefore

$$\mathbf{1}\left\{X = x,\ 0 \le \frac{A-\delta}{h} \le 1\right\} = \mathbf{1}_{E_x(\delta,h)}(O).$$

Integrating this identity gives

$$\int \mathbf{1}\left\{X = x,\ 0 \le \frac{A-\delta}{h} \le 1\right\} dP = P(E_x(\delta,h)).$$

The window-support assumption gives $[\delta, \delta+h] \subseteq [0,1]$. Applying the conditional-density law in stratum x to the Borel set $[\delta, \delta+h]$ yields

$$P(E_x(\delta,h)) = p_x \int_{[\delta,\delta+h]} \pi_x(a)\, da = p_x \int_{\delta}^{\delta+h} \pi_x(a)\, da.$$

The asserted identity follows from the two displayed evaluations of the local window weight. □

Lemma 13 integrates the squared-weight scale over the random design. Under the balance equation, the average stochastic radius has the same h^β order as the deterministic local-polynomial bias term.

The following two lemmas are the finite-sample ingredients for the stabilized procedures. They combine bounded-outcome concentration, deterministic fallback branches, and the local-design controls above.

Lemma 5. [`lem:local-gram-entry-integral`] (Local Gram integral identity) *Fix a law $P \in \mathcal{M}$ with declared constants $(\beta, \kappa, L, c_-, c_+, p_{\min})$, as in Definition 5 and Assumptions 2 to 4. Let $\ell \in \mathbb{N}$, $x \in \mathcal{X}$, and $\delta, h \in \mathbb{R}$. Assume:*

- ***(Constant signs.)*** *$\kappa \ge 0$, $c_+ \ge 0$, and $p_{\min} > 0$.*
- ***(Window.)*** *$h > 0$ and $[\delta, \delta+h] \subseteq [0,1]$.*
- ***(Entries.)*** *$j, k \in \{0, \dots, \ell\}$.*

For $u_\delta(a) = (a-\delta)/h$, write $v_\ell(u) = (1, u, \dots, u^\ell)$ as in Definition 12, and define

$$\phi_j(O) = \mathbf{1}\{0 \le u_\delta(A) \le 1\}\, v_\ell(u_\delta(A))_j.$$

Then

$$\int \mathbf{1}\{X = x,\ 0 \le u_\delta(A) \le 1\}\, \phi_j(O)\phi_k(O)\, dP(O) = p_x \int_\delta^{\delta+h} \pi_x(a)\, v_\ell(u_\delta(a))_j\, v_\ell(u_\delta(a))_k\, da.$$

Proof of Lemma 5. Write

$$T = [\delta, \delta+h], \qquad u_\delta(a) = \frac{a-\delta}{h}.$$

Since $h > 0$, for every real a,

$$0 \le u_\delta(a) \le 1 \iff a \in T.$$

Define the bounded measurable function

$$f(a) = \mathbf{1}\{0 \le u_\delta(a) \le 1\}\, v_\ell(u_\delta(a))_j\, v_\ell(u_\delta(a))_k.$$

For $u \in [0,1]$, the entries $v_\ell(u)_j$ and $v_\ell(u)_k$ are powers of u, so their absolute values are at most one. Hence

$$|f(a)| \le 1 \qquad \text{for every } a.$$

The signs $\kappa \ge 0$ and $c_+ \ge 0$, together with the density envelope on $[0,1]$, give the pointwise domination used for integrability: for Lebesgue-a.e. $a \in [0,1]$,

$$0 \le \pi_x(a) \le c_+ a^\kappa \le c_+.$$

The last inequality follows from $0 \le a \le 1$ and $\kappa \ge 0$, and it is preserved after multiplication by $c_+ \ge 0$. Hence

$$\int_{[0,1]} \pi_x(a)\,da \le c_+ < \infty.$$

Consequently, if b is measurable and $|b(a)| \le 1$, then

$$\int_{[0,1]} |b(a)|\pi_x(a)\,da \le \int_{[0,1]} \pi_x(a)\,da < \infty,$$

so bounded measurable multiples are integrable for the conditional treatment measure in stratum x. The hypothesis $p_{\min} > 0$ is carried with the model hypotheses used below for the stratum identity.

The stratum-restricted treatment law therefore has the following integral form: for every bounded measurable $g : [0,1] \to \mathbb{R}$,

$$\int \mathbf{1}\{X = x\} g(A)\,dP = p_x \int_{[0,1]} g(a)\pi_x(a)\,da.$$

To see this, apply the conditional-density condition in Assumption 2 first to indicators $g = \mathbf{1}_B$, where $B \subseteq [0,1]$ is Borel:

$$\int \mathbf{1}\{X = x\}\mathbf{1}_B(A)\,dP = P(X = x, A \in B) = p_x \int_B \pi_x(a)\,da.$$

Linearity gives the identity for nonnegative simple functions, monotone convergence gives it for nonnegative measurable functions, and decomposition into positive and negative parts gives it for bounded measurable g; the integrability just established makes both sides finite.

Applying this stratum identity to the bounded measurable function f gives

$$\int \mathbf{1}\{X = x\} f(A)\,dP = p_x \int_{[0,1]} f(a)\pi_x(a)\,da.$$

By the definition of ϕ_j and ϕ_k,

$$\mathbf{1}\{X = x,\ 0 \le u_\delta(A) \le 1\}\phi_j(O)\phi_k(O) = \mathbf{1}\{X = x\} f(A).$$

Finally, since $T \subseteq [0,1]$ and $f(a) = 0$ off T, while $f(a) = v_\ell(u_\delta(a))_j v_\ell(u_\delta(a))_k$ on T,

$$\int_{[0,1]} f(a)\pi_x(a)\,da = \int_\delta^{\delta+h} \pi_x(a)\, v_\ell(u_\delta(a))_j\, v_\ell(u_\delta(a))_k\,da.$$

Combining the displayed stratum decomposition with the preceding indicator rewriting yields

$$\int \mathbf{1}\{X = x,\ 0 \le u_\delta(A) \le 1\}\,\phi_j(O)\phi_k(O)\,dP(O) = p_x \int_\delta^{\delta+h} \pi_x(a)\, v_\ell(u_\delta(a))_j\, v_\ell(u_\delta(a))_k\,da.$$

□

Lemma 14 gives the finite-sample coverage statement for the interval in Algorithm 2. The interval combines split-block concentration with a bias-aware local radius; on singular local designs, the bounded fallback radius covers the whole possible threshold-mean contribution in the affected stratum.

Lemma 6. [`lem:scaled-window-coercivity`] (Scaled window coercivity) *Let $\ell \in \mathbb{N}$, and let $\lambda_\star = \lambda_\star(\ell, \kappa, c_-, c_+)$ be the uniform population Gram lower-bound constant from Definition 12, formed with the lower and upper thinning constants in Assumption 4. Suppose that*

- ***(Shape constants.)*** *$\kappa \geq 0$, $c_- > 0$, and $c_+ > 0$.*
- ***(Window.)*** *$\delta \geq 0$ and $h > 0$.*
- ***(Coefficient vector.)*** *$v = (v_0, \ldots, v_\ell) \in \mathbb{R}^{\ell+1}$.*

Then

$$\lambda_\star c_+ \left(\int_\delta^{\delta+h} a^\kappa \, \mathrm{d}a \right) \sum_{j=0}^{\ell} v_j^2 \leq c_- \int_\delta^{\delta+h} \left\{ \sum_{j=0}^{\ell} v_j \left(\frac{a-\delta}{h} \right)^j \right\}^2 a^\kappa \, \mathrm{d}a.$$

Proof of Lemma 6. Put

$$\rho = \frac{\delta}{h}, \qquad D = \int_0^1 (\rho + u)^\kappa \, du, \qquad N = \int_0^1 \left(\sum_{j=0}^{\ell} v_j u^j \right)^2 (\rho + u)^\kappa \, du.$$

Since $\delta \geq 0$ and $h > 0$, one has $\rho \geq 0$. The function $u \mapsto (\rho + u)^\kappa$ is continuous on $[0, 1]$, is nonnegative there, and has value $(\rho + 1)^\kappa > 0$ at $u = 1$. Hence it is positive on a neighborhood of 1, and the interval-integral positivity criterion gives $D > 0$.

By the definition of $M_{\ell,\kappa,\rho}$ in Definition 12, expanding the quadratic form gives

$$v^\top M_{\ell,\kappa,\rho} v = \frac{N}{D}.$$

The uniform coercivity in Lemma 1 supplies the lower boundedness used to compare the defining infimum in $\lambda_\star$ with the ρ-term. The usual rescaling from arbitrary vectors to the Euclidean unit sphere then gives

$$\lambda_\star \sum_{j=0}^{\ell} v_j^2 \leq \frac{c_-}{c_+} \frac{N}{D}.$$

Indeed, for $v = 0$ this is immediate, while for $v \neq 0$ applying the unit-sphere minimum to $v / \{\sum_j v_j^2\}^{1/2}$ yields

$$\inf_{\rho' \geq 0} \lambda_{\min}(M_{\ell,\kappa,\rho'}) \sum_{j=0}^{\ell} v_j^2 \leq v^\top M_{\ell,\kappa,\rho} v.$$

Multiplying by the nonnegative factor c_-/c_+ and using the displayed identity for $v^\top M_{\ell,\kappa,\rho} v$ gives the bound $\lambda_\star \sum_{j=0}^{\ell} v_j^2 \leq (c_-/c_+) \, v^\top M_{\ell,\kappa,\rho} v$. Since $c_+ D > 0$, multiplying by $c_+ D$ gives

$$\lambda_\star c_+ D \sum_{j=0}^{\ell} v_j^2 \leq c_- N.$$

It remains to transport this inequality from $[0,1]$ to the window. With the change of variables $a=\delta+hu$,

$$\int_\delta^{\delta+h} a^\kappa\,da = h\int_0^1 (\delta+hu)^\kappa\,du = h\,h^\kappa\int_0^1(\rho+u)^\kappa\,du = h\,h^\kappa D,$$

because $\delta+hu=h(\rho+u)$ and $h>0$. The same substitution gives

$$\int_\delta^{\delta+h}\left(\sum_{j=0}^{\ell} v_j\left(\frac{a-\delta}{h}\right)^j\right)^2 a^\kappa\,da = h\,h^\kappa N.$$

Since $h\,h^\kappa\ge 0$, multiplying the unit-window inequality by this factor and using the two displayed identities gives exactly

$$\lambda_\star c_+\left(\int_\delta^{\delta+h} a^\kappa\,da\right)\sum_{j=0}^{\ell} v_j^2 \le c_-\int_\delta^{\delta+h}\left(\sum_{j=0}^{\ell} v_j\left(\frac{a-\delta}{h}\right)^j\right)^2 a^\kappa\,da.$$

□

Lemma 15 supplies the estimator side of the upper bound in Theorem 1. The retained-course and atom-mass empirical terms contribute the $n^{-1/2}$ scale, while Lemma 13 aligns the threshold-regression stochastic scale with the Hölder bias scale after multiplication by the collapsed mass.

The phase analysis rests on a separate bandwidth comparison. The result below identifies the edge scale where the boundary form of the local design transitions to the interior form.

Lemma 7. [`lem:intercept-weight-formula`] (Intercept weight formula) *Fix the local-design notation of Definition 12 and the split-block construction of Algorithm 1. Let $B_n=(I_0,I_1,I_2)$ be a deterministic three-way split with pairwise disjoint blocks, each of cardinality at least $\lfloor n/4\rfloor$. Fix a realized sample $z=(O_1,\dots,O_n)$, a stratum $x\in\mathcal{X}$, a degree ℓ, constants κ,c_-,c_+, a threshold δ, a window width h, and an index i.*

Assume:

- ***(Good design.)*** *The good-design event Ω_x holds for $B_n,z,x,\ell,\kappa,c_-,c_+,\delta,h$.*
- ***(Regression block.)*** *The index satisfies $i\in I_2$.*
- ***(Stratum match.)*** *The realized stratum satisfies $X_i=x$.*
- ***(Local window.)*** *With $U_i=(A_i-\delta)/h$, one has $U_i\in[0,1]$.*

Then the exact local-polynomial intercept weight satisfies

$$w_{ix}=e_0^\top G_x^{-1} v_\ell(U_i).$$

Proof of Lemma 7. Let

$$K_i=\mathbf{1}\{i\in I_2,\ X_i=x,\ 0\le U_i\le 1\},\qquad U_i=\frac{A_i-\delta}{h}.$$

On Ω_x, the totalized definition of the exact intercept weight uses the local-polynomial equivalent kernel:

$$w_{ix} = \sum_{k=0}^{\ell} (G_x^{-1})_{0k}\, K_i\, U_i^k.$$

The assumptions $i \in I_2$, $X_i = x$, and $U_i \in [0,1]$ give $K_i = 1$. Hence

$$w_{ix} = \sum_{k=0}^{\ell} (G_x^{-1})_{0k}\, U_i^k = \big(G_x^{-1} v_\ell(U_i)\big)_0 = e_0^\top G_x^{-1} v_\ell(U_i).$$

□

Lemma 16 is the algebraic input behind the phase diagram in Theorem 3. It shows that the bandwidth is comparable to the boundary scale near the edge and to the usual thinned interior bandwidth when the threshold is asymptotically beyond that edge.

The total-Gram probability bound depends on population coercivity and elementary local-window identities. Lemmas 1 to 6 isolate these deterministic and integral facts before they are assembled in Lemma 17.

Lemma 8. [`lem:good-gram-reproduction`] (Good Gram reproduction) *Fix $J, n, \ell \in \mathbb{N}$, a realized sample $z = (O_1, \ldots, O_n)$, a stratum $x \in \mathcal{X}$, constants $\kappa, c_-, c_+, \delta, h \in \mathbb{R}$, and deterministic split blocks $B_n = (I_0, I_1, I_2)$ as in Algorithm 1. Use the local-design notation U_i, N_x, G_x, $\lambda_\star$, Ω_x, and w_{ix} from Definition 12.*

Assume that $\lambda_\star > 0$ and that the good-design event Ω_x holds for $B_n, z, x, \ell, \kappa, c_-, c_+, \delta, h$. Then the exact local-polynomial intercept weights reproduce every monomial coordinate at the threshold: for every $j = 0, \ldots, \ell$,

$$\sum_{i \in I_2} w_{ix} U_i^j = \begin{cases} 1, & j = 0, \\ 0, & j \neq 0. \end{cases}$$

Proof of Lemma 8. Write

$$K_i = \mathbf{1}\{i \in I_2,\ X_i = x,\ 0 \le U_i \le 1\}, \qquad G_x = \sum_i K_i\, v_\ell(U_i) v_\ell(U_i)^\top,$$

where the sum over all sample indices is the same as the sum over I_2, since $K_i = 0$ outside I_2. Set

$$N = N_x, \qquad a = \frac{\lambda_\star N}{2}.$$

On Ω_x, $N > 0$, and the assumed $\lambda_\star > 0$ gives $a > 0$. The good-design inequality is

$$a \sum_{r=0}^{\ell} b_r^2 \le b^\top G_x b \qquad \text{for every } b \in \mathbb{R}^{\ell+1}.$$

If $G_x b = 0$, then the right side is zero, so $\sum_r b_r^2 = 0$ and $b = 0$. Thus G_x is invertible.

On Ω_x, the weights are the equivalent-kernel weights,

$$w_{ix} = K_i\, e_0^\top G_x^{-1} v_\ell(U_i).$$

Consequently, for $0 \le j \le \ell$,

$$\sum_{i \in I_2} w_{ix} U_i^j = e_0^\top G_x^{-1} \left\{ \sum_i K_i \, v_\ell(U_i) U_i^j \right\}.$$

Let e_j be the j-th coordinate vector. Since $v_\ell(U_i)^\top e_j = U_i^j$,

$$\sum_i K_i \, v_\ell(U_i) U_i^j = \sum_i K_i \, v_\ell(U_i) v_\ell(U_i)^\top e_j = G_x e_j.$$

Therefore

$$\sum_{i \in I_2} w_{ix} U_i^j = e_0^\top G_x^{-1} G_x e_j = e_0^\top e_j = \begin{cases} 1, & j = 0, \\ 0, & j \neq 0. \end{cases}$$

□

Lemma 1 gives a uniform lower eigenvalue bound for the shifted reference moment matrices. The bound is uniform in the shift parameter, which is what permits thresholds ranging from the boundary region to fixed positive interior points.

Lemma 9. [`lem:intercept-weight-pointwise`] (Pointwise intercept-weight bound) *Fix integers J, n, ℓ, a split $B_n = (I_0, I_1, I_2)$, a realized sample $z = (O_1, \dots, O_n)$, a stratum $x \in \mathcal{X}$, constants $\kappa, c_-, c_+, \delta, h \in \mathbb{R}$, and an index $i \in \{1, \dots, n\}$. With N_x, $\lambda_\star$, Ω_x, and w_{ix} as in Definition 12, assume:*

- ***(Split.)*** *The split B_n has three pairwise disjoint deterministic blocks I_0, I_1, I_2, each of cardinality at least $\lfloor n/4 \rfloor$, as in Algorithm 1.*
- ***(Positive population Gram constant.)*** *$\lambda_\star > 0$.*
- ***(Good design.)*** *The good-design event Ω_x holds at the realized sample z.*

Then the exact local-polynomial intercept weight at index i satisfies

$$|w_{ix}| \le \left(\frac{\lambda_\star N_x}{2} \right)^{-1} \sqrt{\ell + 1}.$$

Proof of Lemma 9. Let

$$N = N_x, \qquad a = \frac{\lambda_\star N}{2}.$$

On Ω_x, $N > 0$; together with $\lambda_\star > 0$, this gives $a > 0$. The good-design inequality states that

$$a \|b\|_2^2 \le b^\top G_x b \qquad \text{for every } b \in \mathbb{R}^{\ell+1}.$$

First suppose that i is active, meaning $i \in I_2$, $X_i = x$, and $U_i \in [0, 1]$. Put

$$s = v_\ell(U_i), \qquad \xi_i = G_x^{-1} s \in \mathbb{R}^{\ell+1}.$$

The coercivity inequality implies the inverse bound

$$\|G_x^{-1} s\|_2 \le a^{-1} \|s\|_2.$$

Indeed, $G_x \xi_i = s$, so
$$a\|\xi_i\|_2^2 \le \xi_i^\top G_x \xi_i = \xi_i^\top s \le \|\xi_i\|_2 \|s\|_2,$$
and division by $\|\xi_i\|_2$ is harmless, with the case $\|\xi_i\|_2 = 0$ immediate.

By Lemma 7,
$$|w_{ix}| = \left|(G_x^{-1} s)_0\right| \le \|G_x^{-1} s\|_2 \le a^{-1}\|s\|_2.$$
Since $U_i \in [0,1]$,
$$\|s\|_2^2 = \sum_{j=0}^{\ell} U_i^{2j} \le \ell + 1,$$
and hence
$$|w_{ix}| \le \left(\frac{\lambda_\star N_x}{2}\right)^{-1} \sqrt{\ell+1}.$$

If i is inactive, then the local kernel factor in the weight is zero on Ω_x, so
$$w_{ix} = 0,$$
and the same bound follows because the right side is nonnegative. □

Lemma 2 verifies the positivity of the Gram constant used in Definition 12. This makes the good-design event meaningful uniformly over the polynomial-thinning envelope.

Lemma 10. [`lem:bandwidth-balance`] (Eventual bandwidth balance) *Let h_n denote the information-balance bandwidth from Definition 8 formed with $\delta_n, \beta, \kappa, \bar{\delta}$. Suppose that:*

- ***(Hölder exponent.)*** *The exponent β appearing in Assumption 5 satisfies $\beta > 0$.*
- ***(Polynomial-thinning exponent.)*** *The exponent κ appearing in Assumption 4 satisfies $\kappa \ge 0$.*
- ***(Threshold cap.)*** $0 < \bar{\delta} < 1$.
- ***(Threshold sequence.)*** *$(\delta_n)_{n\ge 1}$ is a deterministic real sequence satisfying $\delta_n \in [0, \bar{\delta}]$ for every n.*

Then, for all sufficiently large n,
$$0 < h_n, \qquad h_n \le 1 - \bar{\delta}, \qquad n h_n^{2\beta+1}(\delta_n + h_n)^\kappa = 1.$$

Proof of Lemma 10. Put $H = 1 - \bar{\delta}$ and $p = 2\beta + 1$. The hypotheses give $H > 0$ and $p > 0$. Since $H^p H^\kappa > 0$, choose N so large that
$$N > \frac{1}{H^p H^\kappa}.$$
Because $1/(H^p H^\kappa) > 0$, this choice gives $N > 0$, and hence every $n \ge N$ satisfies $n > 0$. For $n \ge N$, define
$$F_n(h) = n h^p (\delta_n + h)^\kappa, \qquad 0 \le h \le H.$$
The threshold condition gives $\delta_n \ge 0$, so
$$F_n(H) = n H^p (\delta_n + H)^\kappa \ge n H^p H^\kappa > 1,$$

because $H \le \delta_n + H$ and $t \mapsto t^\kappa$ is nondecreasing on $[0, \infty)$. For continuity on $[0, H]$, use the real-power convention $0^0 = 1$. If $\kappa = 0$, then $(\delta_n + h)^\kappa = 1$ on the whole interval and $F_n(h) = nh^p$, which is continuous. If $\kappa \ne 0$, then $\kappa > 0$, and the map $h \mapsto (\delta_n + h)^\kappa$ is continuous on $[0, H]$ because $\delta_n + h \ge 0$; multiplying by the continuous factor nh^p gives continuity of F_n. In both cases

$$F_n(0) = n\, 0^p (\delta_n + 0)^\kappa = 0,$$

since $p > 0$. Continuity of F_n on $[0, H]$ therefore gives a point $\eta_n \in [0, H]$ with

$$F_n(\eta_n) = 1.$$

This point is strictly positive, since $F_n(0) = 0$.

The same monotonicity gives strict increase of F_n on $[0, H]$: if $0 \le a < b \le H$, then $a^p < b^p$, while

$$(\delta_n + a)^\kappa \le (\delta_n + b)^\kappa.$$

Moreover $n > 0$ by the preceding choice of N, and $\delta_n + b > 0$, so

$$n(\delta_n + b)^\kappa > 0.$$

Multiplying the strict inequality for a^p and the weak inequality for the second power by this positive factor gives $F_n(a) < F_n(b)$. Consequently the crossing set in Definition 8 is exactly

$$\{h \in (0, H] : F_n(h) \ge 1\} = [\eta_n, H].$$

Its infimum is η_n, so $h_n = \eta_n$. Thus, for every $n \ge N$,

$$0 < h_n, \qquad h_n \le 1 - \bar{\delta}, \qquad n h_n^{2\beta+1} (\delta_n + h_n)^\kappa = 1.$$

□

Lemma 3 gives a simple lower bound on local design mass under the power envelope. It is used to compare the realized local count with $h(\delta + h)^\kappa$.

Lemma 11. [`lem:local-gram-coercivity`] (Local Gram coercivity) *Fix an observed law $P \in \mathcal{M}$ as in Definition 5, with conditional-density, stratum-mass, and polynomial-thinning components as in Assumptions 2 to 4. Fix constants $\beta, \kappa, L, c_-, c_+$, and $p_{\min}$, and a degree $\ell \in \mathbb{N}$. Suppose that*

- ***(Design constants.)*** *$\kappa \ge 0$, $c_- > 0$, $c_+ > 0$, and $p_{\min} > 0$.*
- ***(Local window.)*** *$x \in \mathcal{X}$, $\delta \ge 0$, $h > 0$, and $\delta + h \le 1$.*

For an observation $O = (X, A, Y)$, set

$$p_x^{\mathrm{loc}}(\delta, h) = \mathbb{E}_P\left[\mathbf{1}\left\{X = x,\ 0 \le \frac{A - \delta}{h} \le 1\right\}\right],$$

and, for $j = 0, \ldots, \ell$, define the local monomial feature

$$\phi_j(O) = \mathbf{1}\left\{0 \le \frac{A - \delta}{h} \le 1\right\} \left(\frac{A - \delta}{h}\right)^j.$$

Then, with $\lambda_\star = \lambda_\star(\ell, \kappa, c_-, c_+)$ denoting the uniform population Gram lower-bound constant from Definition 12, every $v = (v_0, \ldots, v_\ell) \in \mathbb{R}^{\ell+1}$ satisfies

$$\lambda_\star\, p_x^{\mathrm{loc}}(\delta, h) \sum_{j=0}^{\ell} v_j^2 \le \sum_{j=0}^{\ell}\sum_{k=0}^{\ell} v_j v_k\, \mathbb{E}_P\left[\mathbf{1}\left\{X = x,\ 0 \le \frac{A-\delta}{h} \le 1\right\}\phi_j(O)\phi_k(O)\right].$$

Proof of Lemma 11. Let

$$T = [\delta, \delta + h], \qquad q_v(a) = \sum_{j=0}^{\ell} v_j \left(\frac{a-\delta}{h}\right)^j, \qquad \phi_j(O) = \mathbf{1}\left\{0 \le \frac{A-\delta}{h} \le 1\right\}\left(\frac{A-\delta}{h}\right)^j.$$

The inequalities $\delta \ge 0$ and $\delta + h \le 1$ give $T \subset [0, 1]$. Since $h > 0$, the condition $0 \le (A-\delta)/h \le 1$ is equivalent to $A \in T$. Hence Lemma 4 gives

$$\mathbb{E}_P\left[\mathbf{1}\left\{X = x,\ 0 \le \frac{A-\delta}{h} \le 1\right\}\right] = p_x \int_T \pi_x(a)\,\mathrm{d}a.$$

The left-hand side is $p_x^{\mathrm{loc}}(\delta, h)$.

For each pair (j, k), Lemma 5 gives the corresponding density-weighted local Gram entry. Summing those identities over j and k, and expanding $q_v(a)^2$, yields

$$\begin{aligned}\sum_{j=0}^{\ell}\sum_{k=0}^{\ell} v_j v_k\, \mathbb{E}_P&\left[\mathbf{1}\left\{X = x,\ 0 \le \frac{A-\delta}{h} \le 1\right\}\phi_j(O)\phi_k(O)\right]\\ &= p_x \int_T q_v(a)^2 \pi_x(a)\,\mathrm{d}a.\end{aligned}$$

The stratum-mass condition in Assumption 3 and $p_{\min} > 0$ give $p_x \ge 0$. Combining the displayed local mass identity for $p_x^{\mathrm{loc}}(\delta, h)$ with the upper envelope in Assumption 4 gives

$$p_x^{\mathrm{loc}}(\delta, h) \le p_x\, c_+ \int_T a^\kappa\,\mathrm{d}a.$$

The lower envelope in Assumption 4, together with $q_v(a)^2 \ge 0$, gives

$$p_x\, c_- \int_T q_v(a)^2 a^\kappa\,\mathrm{d}a \le p_x \int_T q_v(a)^2 \pi_x(a)\,\mathrm{d}a.$$

By Lemma 2, the assumptions $\kappa \ge 0$, $c_- > 0$, and $c_+ > 0$ imply $\lambda_\star > 0$; in particular $\lambda_\star \ge 0$. Applying Lemma 6 with this δ, h and coefficient vector v gives

$$\lambda_\star c_+ \left(\int_T a^\kappa\,\mathrm{d}a\right) \sum_{j=0}^{\ell} v_j^2 \le c_- \int_T q_v(a)^2 a^\kappa\,\mathrm{d}a.$$

Combining the mass upper bound with $\lambda_\star \ge 0$ and $\sum_j v_j^2 \ge 0$, then applying the scaled-window coercivity bound and the lower density-envelope bound, gives

$$\begin{aligned}\lambda_\star\, p_x^{\mathrm{loc}}(\delta, h) \sum_{j=0}^{\ell} v_j^2 &\le p_x\, \lambda_\star c_+ \left(\int_T a^\kappa\,\mathrm{d}a\right) \sum_{j=0}^{\ell} v_j^2\\ &\le p_x\, c_- \int_T q_v(a)^2 a^\kappa\,\mathrm{d}a\\ &\le p_x \int_T q_v(a)^2 \pi_x(a)\,\mathrm{d}a.\end{aligned}$$

Using the Gram identity above for the final term proves

$$\lambda_\star\, p_x^{\mathrm{loc}}(\delta,h)\sum_{j=0}^{\ell} v_j^2 \le \sum_{j=0}^{\ell}\sum_{k=0}^{\ell} v_j v_k\,\mathbb{E}_P\left[\mathbf{1}\left\{X=x,\ 0\le \frac{A-\delta}{h}\le 1\right\}\phi_j(O)\phi_k(O)\right].$$

□

Lemma 4 expresses the local-window probability through the stratum mass and conditional density. It is the population counterpart of the count N_x in Definition 12.

Lemma 12. [`lem:good-gram-weights`] (Good Gram weights) *Fix $J, n, \ell \in \mathbb{N}$, a split $B_n = (I_0, I_1, I_2)$ with pairwise disjoint blocks satisfying $|I_r| \ge \lfloor n/4 \rfloor$ for $r = 0, 1, 2$, an observed sample $z = (O_1, \dots, O_n)$, and a stratum $x \in \mathcal{X}$. Let $\kappa, c_-, c_+, \delta, h \in \mathbb{R}$. Suppose:*

- ***(Positive bandwidth.)*** *$h > 0$.*
- ***(Positive population Gram constant.)*** *The constant $\lambda_\star$ from Definition 12 satisfies $\lambda_\star > 0$.*
- ***(Good local design.)*** *The good-design event Ω_x from Definition 12 holds for $B_n, z, x, \ell, \kappa, c_-, c_+, \delta, h$.*

Then the exact local-polynomial intercept weights w_{ix} and rescaled coordinates U_i from Definition 12 satisfy, for every $j = 0, \dots, \ell$,

$$\sum_{i\in I_2} w_{ix} U_i^j = \mathbf{1}\{j=0\}.$$

They also obey the two deterministic weight bounds

$$\sum_{i\in I_2} |w_{ix}| \le \frac{2\sqrt{\ell+1}}{\lambda_\star},$$

and, with N_x denoting the local count in Definition 12,

$$\sum_{i\in I_2} w_{ix}^2 \le \frac{4(\ell+1)}{\lambda_\star^2 N_x}.$$

Proof of Lemma 12. Write

$$\lambda = \lambda_\star(\ell,\kappa,c_-,c_+), \qquad N = N_x, \qquad G = G_x.$$

On Ω_x, $\lambda > 0$, $N > 0$, and

$$v^\top G v \ge \frac{\lambda N}{2}\sum_{r=0}^{\ell} v_r^2 \qquad \text{for every } v \in \mathbb{R}^{\ell+1}.$$

Thus G is nonsingular: if $Gv = 0$, the coercivity bound $v^\top G v \ge (\lambda N/2)\sum_r v_r^2$ with $\lambda N > 0$ forces $\sum_r v_r^2 = 0$, hence $v = 0$. The resulting monomial identity is the good-Gram reproduction statement in Lemma 8.

For active observations $i \in I_2$ with $X_i = x$ and $0 \le U_i \le 1$, the weight is

$$w_{ix} = e_0^\top G^{-1} v_\ell(U_i),$$

and all inactive terms in the sum over I_2 have weight 0. Therefore, for $j = 0, \ldots, \ell$,

$$\begin{aligned}\sum_{i \in I_2} w_{ix} U_i^j &= e_0^\top G^{-1} \sum_{i \in I_2} \mathbf{1}\{X_i = x,\ 0 \le U_i \le 1\} v_\ell(U_i) U_i^j \\ &= e_0^\top G^{-1} G e_j = \mathbf{1}\{j = 0\}.\end{aligned}$$

Next let $a = \lambda N/2$. The pointwise conclusion to be used is Lemma 9. For an active i, set $s_i = v_\ell(U_i)$ and $z_i = G^{-1} s_i$. Coercivity and Cauchy–Schwarz give

$$a\|z_i\|_2^2 \le z_i^\top G z_i = z_i^\top s_i \le \|z_i\|_2 \|s_i\|_2,$$

hence

$$\|G^{-1} s_i\|_2 \le a^{-1} \|s_i\|_2.$$

Since $0 \le U_i \le 1$,

$$\|s_i\|_2^2 = \sum_{r=0}^{\ell} U_i^{2r} \le \ell + 1,$$

and therefore

$$|w_{ix}| \le \left(\frac{\lambda N}{2}\right)^{-1} \sqrt{\ell + 1}$$

for every active i, while inactive weights vanish.

The number of active indices is N. Summing the pointwise bound gives

$$\sum_{i \in I_2} |w_{ix}| \le N \left(\frac{\lambda N}{2}\right)^{-1} \sqrt{\ell + 1} = \frac{2\sqrt{\ell + 1}}{\lambda}.$$

Squaring the same pointwise bound and summing gives

$$\sum_{i \in I_2} w_{ix}^2 \le N \left(\frac{\lambda N}{2}\right)^{-2} (\ell + 1) = \frac{4(\ell + 1)}{\lambda^2 N}.$$

□

Lemma 5 identifies each population Gram entry with a one-dimensional integral over the local treatment window. Combined with the density envelope, it connects the empirical Gram matrix in Definition 12 to the shifted reference matrices.

Lemma 13. [`lem:weight-energy-balance`] (Balanced weight energy) *Let $J, n, \ell \in \mathbb{N}$, let P be an observed-data law on J strata, let $B_n = (I_0, I_1, I_2)$ be a deterministic split, and fix a stratum $x \in \mathcal{X}$. Let $\beta, \kappa, L, c_-, c_+, p_{\min}, \delta, h \in \mathbb{R}$, and define the observed product sampling law by*

$$P^{\otimes n} = \bigotimes_{j=1}^{n} P.$$

Suppose that:

- ***(Model and sampling.)*** *P belongs to the observed model class $\mathcal{M}$ in Definition 5 with fixed constants $\beta, \kappa, L, c_-, c_+, p_{\min}$, including the polynomial-thinning, stratum-mass, and Hölder-regression components in Assumptions 3 to 5; the observed sample is independently drawn from P as in Assumption 1; and B_n is a three-way split as in Algorithm 1.*

- ***(Numerical conditions.)*** *$n \geq 8$, $\beta > 0$, $\kappa \geq 0$, $c_- > 0$, $c_+ \geq 0$, $p_{\min} > 0$, and $\delta \geq 0$.*

- ***(Bandwidth and design.)*** *$h > 0$, $\delta + h \leq 1$, and the uniform population Gram lower-bound constant $\lambda_\star = \lambda_\star(\ell, \kappa, c_-, c_+)$ from Definition 12 satisfies $\lambda_\star > 0$.*

- ***(Balance.)*** *The bandwidth is information-balanced:*

$$nh^{2\beta+1}(\delta + h)^\kappa = 1.$$

For the exact local-polynomial intercept weights w_{ix} from Definition 12,

$$\int \left(\sum_{i \in I_2} w_{ix}(z)^2 \right)^{1/2} dP^{\otimes n}(z) \leq \left(\frac{336\{4(\ell+1)/\lambda_\star^2\}}{p_{\min} c_- / (2 \cdot 2^\kappa)} \right)^{1/2} h^\beta.$$

Proof of Lemma 13. Set

$$p = \mathbb{E}_P\left[\mathbf{1}\left\{ X = x,\ 0 \leq \frac{A - \delta}{h} \leq 1 \right\} \right], \qquad D = \frac{4(\ell+1)}{\lambda_\star^2}, \qquad K = \frac{p_{\min} c_-}{2 \cdot 2^\kappa}.$$

Because $h > 0$, the event in the definition of p is $\{X = x,\ A \in [\delta, \delta + h]\}$. Since $\delta \geq 0$ and $\delta + h \leq 1$, this window is contained in $[0, 1]$. The conditional-density, stratum-mass, and polynomial-thinning components of Definition 5, equivalently Assumptions 2 to 4, give

$$\begin{aligned} p &= p_x \int_{[\delta, \delta+h]} \pi_x(a)\, da \\ &\geq p_{\min} c_- \int_\delta^{\delta+h} a^\kappa\, da. \end{aligned}$$

For $a \in [\delta + h/2, \delta + h]$, the assumptions $\delta \geq 0$ and $\kappa \geq 0$ imply $a^\kappa \geq ((\delta + h)/2)^\kappa$. Hence

$$\int_\delta^{\delta+h} a^\kappa\, da \geq \int_{\delta+h/2}^{\delta+h} \left(\frac{\delta + h}{2} \right)^\kappa da = \left(\frac{\delta + h}{2} \right)^\kappa \frac{h}{2}.$$

Thus

$$p_{\min} c_- \left(\frac{\delta + h}{2} \right)^\kappa \frac{h}{2} \leq p.$$

Since $\delta + h > 0$,

$$\left(\frac{\delta + h}{2} \right)^\kappa = \frac{(\delta + h)^\kappa}{2^\kappa},$$

and therefore

$$K\, h(\delta + h)^\kappa \leq p.$$

The constants satisfy $K > 0$, $h > 0$, and $(\delta + h)^\kappa > 0$, so $p > 0$.

Let $m = |I_2|$. The admissible split-size condition in Definition 1 gives $|I_2| \geq \lfloor n/4 \rfloor$, and together with $n \geq 8$ this yields

$$\frac{n}{8} \leq m.$$

Thus $m > 0$, and $p > 0$ gives $mp > 0$. Let $N_x(z)$ be the local count in Definition 12, evaluated for the displayed (δ, h). The count decomposition used for the weight energy is

$$\sum_{i \in I_2} w_{ix}(z)^2 \leq \frac{2D}{mp} + D\,\mathbf{1}\Big\{N_x(z) \leq \frac{mp}{2}\Big\}.$$

Indeed, on Ω_x, Lemma 12 gives

$$\sum_{i \in I_2} w_{ix}(z)^2 \leq \frac{D}{N_x(z)};$$

on Ω_x^c, the weights in Definition 12 are zero. Consequently the energy is at most D on the low-count event $N_x(z) \leq mp/2$, while on its complement the good-design case gives $D/N_x(z) \leq 2D/(mp)$ and the Ω_x^c case gives zero. Under Definition 2 and Assumption 1, the variables defining N_x on the third split block are independent Bernoulli variables with mean p. The Bernoulli lower-tail bound gives

$$P^{\otimes n}\Big\{z : N_x(z) \leq \frac{mp}{2}\Big\} \leq 2\exp\Big(-\frac{mp}{20}\Big).$$

Integrating the pointwise decomposition gives

$$\begin{aligned}\int \sum_{i \in I_2} w_{ix}(z)^2\, dP^{\otimes n}(z) &\leq \frac{2D}{mp} + D\,P^{\otimes n}\Big\{z : N_x(z) \leq \frac{mp}{2}\Big\} \\ &\leq \frac{2D}{mp} + 2D\exp\Big(-\frac{mp}{20}\Big).\end{aligned}$$

Jensen's inequality for the square root therefore gives

$$\int \left(\sum_{i \in I_2} w_{ix}(z)^2\right)^{1/2} dP^{\otimes n}(z) \leq \left\{\frac{2D}{mp} + 2D\exp\Big(-\frac{mp}{20}\Big)\right\}^{1/2}.$$

The balance condition rewrites the effective sample size as

$$n\,h(\delta + h)^{\kappa} = h^{-2\beta}.$$

Together with $m \geq n/8$ and $p \geq Kh(\delta + h)^{\kappa}$, this gives

$$mp \geq \frac{K}{8}h^{-2\beta}, \qquad (mp)^{-1} \leq \frac{8}{K}h^{2\beta}.$$

Also $\exp(-t/20) \leq 20/t$ for $t > 0$, applied with $t = mp$. Therefore

$$\begin{aligned}\frac{2D}{mp} + 2D\exp\Big(-\frac{mp}{20}\Big) &\leq 42D(mp)^{-1} \\ &\leq \frac{336D}{K}h^{2\beta}.\end{aligned}$$

Taking square roots and using $h > 0$ yields

$$\left\{\frac{2D}{mp} + 2D\exp\Big(-\frac{mp}{20}\Big)\right\}^{1/2} \le \left(\frac{336D}{K}\right)^{1/2} h^{\beta}.$$

Substituting the definitions of D and K gives exactly

$$\int \left(\sum_{i\in I_2} w_{ix}(z)^2\right)^{1/2} dP^{\otimes n}(z) \le \left(\frac{336\{4(\ell+1)/\lambda_\star^2\}}{p_{\min}c_-/(2\cdot 2^{\kappa})}\right)^{1/2} h^{\beta}.$$

□

Lemma 6 converts the normalized shifted-moment lower bound into the physical treatment scale. This is the population inequality used before applying concentration to the realized Gram matrix.

Lemmas 7 to 9 record pointwise properties of the exact intercept weights on the good-design event. They are deterministic consequences of the total Gram matrix and the affine reproduction equations.

Lemma 14. [`lem:stabilized-interval-honesty`] (Stabilized interval honesty) *Fix integers J and n, deterministic split blocks $B = (I_0, I_1, I_2)$, constants $\beta, \kappa, L, c_-, c_+, p_{\min}, \bar\delta, \alpha$, and a threshold δ. Let h_n be the information-balance bandwidth from Definition 8 evaluated at $(n, \delta, \beta, \kappa, \bar\delta)$, and let CI_n be the stabilized interval from Algorithm 2 formed with split B, bandwidth h_n, and threshold δ. Suppose that:*

- ***(Split blocks.)*** *$B = (I_0, I_1, I_2)$ consists of three deterministic, pairwise-disjoint subsets of $\{1, \dots, n\}$ with*

$$|I_0| \ge \lfloor n/4 \rfloor, \qquad |I_1| \ge \lfloor n/4 \rfloor, \qquad |I_2| \ge \lfloor n/4 \rfloor.$$

- ***(Regime constants.)*** *The declared constants satisfy*

$$J \ge 1, \quad \beta > 0, \quad \kappa \ge 0, \quad L > 0, \quad 0 < c_- \le \kappa{+}1 \le c_+, \quad 0 < p_{\min} \le 1/J, \quad 0 < \bar\delta < 1, \quad 0 < \alpha < 1/2.$$

- ***(Sample size.)*** *$n \ge 4$.*
- ***(Threshold range.)*** *$\delta \in [0, \bar\delta]$.*
- ***(Interior bandwidth.)*** *$0 < h_n$ and $\delta + h_n \le 1$.*

Then CI_n is an observed-sample measurable interval, and for every observed-data law P in the clamp model $\mathcal{M}$ of Definition 5,

$$1 - \alpha \le P^{\otimes n}\{\theta_\delta(P) \in \mathrm{CI}_n(O_1, \dots, O_n)\},$$

where $\theta_\delta(P)$ is the unsmoothed observed-data clamp target from Definition 7.

Proof of Lemma 14. Write $h = h_n$, and let

$$t_\alpha = \sqrt{\frac{\log(12J/\alpha)}{2}}, \qquad b_0 = \frac{t_\alpha}{\sqrt{|I_0|}}, \qquad b_1 = \frac{t_\alpha}{\sqrt{|I_1|}}.$$

The regime assumptions imply $t_\alpha > 0$, and the split-size and sample-size assumptions imply $|I_0|, |I_1| > 0$. The center $\widehat{\theta}_n$ in Algorithm 1 and every stratum contribution D_x in Algorithm 2 are measurable functions of the observed sample. Hence the endpoint maps

$$z \mapsto \max\left\{0, \widehat{\theta}_n(z) - b_0 - \sum_x D_x(z)\right\}, \qquad z \mapsto \min\left\{1, \widehat{\theta}_n(z) + b_0 + \sum_x D_x(z)\right\}$$

are measurable. The radius $b_0 + \sum_x D_x$ is nonnegative and $\widehat{\theta}_n(z) \in [0,1]$, so these endpoints are ordered and define an observed-sample measurable interval.

Fix $P \in \mathcal{M}$. Define

$$f_0(O) = Y\mathbf{1}\{A > \delta\}, \qquad f_x(O) = \mathbf{1}\{X = x,\ A \le \delta\}.$$

By Definition 7 and the conditional-density component of Definition 5,

$$\mathbb{E}_P f_0(O) = \nu_\delta(P), \qquad \mathbb{E}_P f_x(O) = p_x q_x(\delta).$$

Let

$$\mathcal{B}_0 = \{|\widehat{\nu}_n - \mathbb{E}_P f_0(O)| > b_0\},$$
$$\mathcal{B}_{1x} = \left\{\left|\widehat{b}_x - P(X = x, A \le \delta)\right| > b_1\right\},$$

and

$$\mathcal{B}_{2x} = \left\{\Omega_x,\ t_\alpha \left(\sum_{i \in I_2} w_{ix}^2\right)^{1/2} \le \left|\sum_{i \in I_2} w_{ix}\{Y_i - \mu^P_{X_i}(A_i)\}\right|\right\}.$$

Hoeffding's inequality on the first two split blocks gives

$$P^{\otimes n}(\mathcal{B}_0) \le \frac{\alpha}{6J}, \qquad P^{\otimes n}(\mathcal{B}_{1x}) \le \frac{\alpha}{6J} \quad \text{for every } x.$$

The self-normalized weighted Hoeffding bound on the third block gives

$$P^{\otimes n}(\mathcal{B}_{2x}) \le 2e^{-2t_\alpha^2} \le \frac{\alpha}{6J} \quad \text{for every } x.$$

Let

$$\mathcal{B} = \mathcal{B}_0 \cup \bigcup_x \mathcal{B}_{1x} \cup \bigcup_x \mathcal{B}_{2x}.$$

Since $J \ge 1$,

$$P^{\otimes n}(\mathcal{B}) \le \frac{\alpha}{6J} + J\frac{\alpha}{6J} + J\frac{\alpha}{6J} \le \alpha.$$

On $\mathcal{B}^c$, the retained-course and atom-mass deviations satisfy

$$|\widehat{\nu}_n - \nu_\delta(P)| \le b_0, \qquad |\widehat{b}_x - p_x q_x(\delta)| \le b_1 \quad \text{for every } x.$$

In the notation of Definition 12, set

$$\ell = \ell(\beta), \qquad \lambda_\star = \lambda_\star(\ell, \kappa, c_-, c_+).$$

The regime assumptions give $\ell \in \mathbb{N}$, $\kappa \geq 0$, $c_- > 0$, and $c_+ > 0$, so Lemma 2 yields

$$\lambda_\star > 0.$$

For a stratum x with Ω_x, the hypotheses $h > 0$, $\lambda_\star > 0$, and Ω_x are therefore in force, and Lemma 12 gives the reproduction identities

$$\sum_{i \in I_2} w_{ix} U_i^j = \mathbf{1}\{j = 0\}, \qquad j = 0, \ldots, \ell.$$

With

$$T_x(a) = \sum_{j=0}^{\ell} \frac{(\mu_x^P)^{(j)}(\delta)}{j!}(a - \delta)^j,$$

these identities imply

$$\sum_{i \in I_2} w_{ix} T_x(A_i) = \mu_x^P(\delta).$$

The inactive-window weights are zero, and on active summands $X_i = x$ and $A_i \in [\delta, \delta{+}h] \subset [0, 1]$. The Hölder condition in Definition 5 therefore yields

$$\left| \sum_{i \in I_2} w_{ix} \mu_{X_i}^P(A_i) - \mu_x^P(\delta) \right| \leq L h^\beta \sum_{i \in I_2} |w_{ix}|.$$

Since $\mu_x^P(\delta) \in [0, 1]$, projection onto $[0, 1]$ is contractive around this target, and the decomposition

$$\sum_{i \in I_2} w_{ix} Y_i - \mu_x^P(\delta) = \sum_{i \in I_2} w_{ix}\{Y_i - \mu_{X_i}^P(A_i)\} + \left(\sum_{i \in I_2} w_{ix} \mu_{X_i}^P(A_i) - \mu_x^P(\delta) \right)$$

gives

$$\left|\widetilde{\mu}_x - \mu_x^P(\delta)\right| \leq L h^\beta \sum_{i \in I_2} |w_{ix}| + \left| \sum_{i \in I_2} w_{ix}\{Y_i - \mu_{X_i}^P(A_i)\} \right|.$$

On $\mathcal{B}^c \cap \Omega_x$, the last absolute value is at most $t_\alpha(\sum_{i \in I_2} w_{ix}^2)^{1/2}$, so

$$\left|\widetilde{\mu}_x - \mu_x^P(\delta)\right| \leq B_x,$$

where B_x is the local regression bias-plus-noise radius in Algorithm 2.

The two displayed deviation bounds imply the deterministic error bound

$$\left|\widehat{\theta}_n - \theta_\delta(P)\right| \leq b_0 + \sum_x D_x.$$

To see the algebra, set $a_x = p_x q_x(\delta)$ and $\widehat{a}_x = \widehat{b}_x$. The atom deviation gives

$$a_x \leq \widehat{a}_x + b_1 \quad \text{and} \quad |\widehat{a}_x - a_x| \leq b_1.$$

Using Algorithm 1 and Definition 7 and contraction of the projection around the target $\theta_\delta(P) \in [0,1]$,

$$\left|\widehat{\theta}_n - \theta_\delta(P)\right| \le |\widehat{\nu}_n - \nu_\delta(P)| + \sum_x \left|\widehat{a}_x\widetilde{\mu}_x - a_x\mu_x^P(\delta)\right|.$$

On Ω_x, the product decomposition

$$\widehat{a}_x\widetilde{\mu}_x - a_x\mu_x^P(\delta) = (\widehat{a}_x - a_x)\widetilde{\mu}_x + a_x\{\widetilde{\mu}_x - \mu_x^P(\delta)\}$$

and the bounds $0 \le \widetilde{\mu}_x \le 1$, $a_x \le \widehat{a}_x + b_1$, and $|\widetilde{\mu}_x - \mu_x^P(\delta)| \le B_x$ give

$$\left|\widehat{a}_x\widetilde{\mu}_x - a_x\mu_x^P(\delta)\right| \le b_1 + (\widehat{b}_x + b_1)B_x = D_x.$$

On Ω_x^c, both $\widetilde{\mu}_x$ and $\mu_x^P(\delta)$ lie in $[0,1]$, so

$$\left|\widehat{a}_x\widetilde{\mu}_x - a_x\mu_x^P(\delta)\right| \le \max\{\widehat{a}_x, a_x\} \le \widehat{b}_x + b_1 = D_x.$$

Together with $|\widehat{\nu}_n - \nu_\delta(P)| \le b_0$, these are exactly the two branches of the interval contribution in Algorithm 2.

Because $\theta_\delta(P) \in [0,1]$, the two displayed branches of the interval contribution place $\theta_\delta(P)$ inside the clipped interval $\mathrm{CI}_n(z)$ for every $z \in \mathcal{B}^c$. Hence

$$P^{\otimes n}\{\theta_\delta(P) \in \mathrm{CI}_n(O_1,\dots,O_n)\} \ge P^{\otimes n}(\mathcal{B}^c) \ge 1-\alpha.$$

Since $P \in \mathcal{M}$ was arbitrary, the same lower bound holds for every law in the stated model class, together with the measurability assertion proved above. □

Lemma 7 gives the explicit intercept-weight representation for observations in the local stratum-window. It is the algebraic formula behind the threshold regression term in Algorithm 1.

Lemma 15. [lem:stabilized-risk-upper] (Stabilized risk upper bound) *Assume that the declared constants satisfy*

- ***(Strata.)*** $J \ge 1$.
- ***(Smoothness and thinning.)*** $\beta > 0$ *and* $\kappa \ge 0$.
- ***(Model radii.)*** $L > 0$, $c_- > 0$, $c_- \le \kappa + 1 \le c_+$, *and* $p_{\min} > 0$ *with* $p_{\min} \le 1/J$.
- ***(Threshold and calibration.)*** $0 < \bar{\delta} < 1$ *and* $0 < \alpha < 1/2$.

Then there is a constant $C_{\mathrm{up}} > 0$, depending on these declared constants, such that the following holds. For every deterministic threshold sequence $(\delta_n)_{n\ge1}$ with $0 \le \delta_n \le \bar{\delta}$ for every n, and for every deterministic split sequence $B_\bullet = (B_n)_{n\ge1}$ in which each $B_n = (I_{0,n}, I_{1,n}, I_{2,n})$ consists of pairwise disjoint subsets of $\{1,\dots,n\}$ with

$$|I_{0,n}| \ge \lfloor n/4 \rfloor, \qquad |I_{1,n}| \ge \lfloor n/4 \rfloor, \qquad |I_{2,n}| \ge \lfloor n/4 \rfloor,$$

there is $N < \infty$ such that, for every $n \ge N$,

$$\sup_{P\in\mathcal{M}} \mathbb{E}_{P^{\otimes n}} \left|\widehat{\theta}^{\mathrm{TG}}_{n,B_n} - \theta_{\delta_n}(P)\right| \le C_{\mathrm{up}}\left(n^{-1/2} + \delta_n^{\kappa+1}h_n^{\beta}\right),$$

where $\mathcal{M}$ is the model class in Definition 5, $\theta_{\delta_n}(P)$ is the clamp target in Definition 7, $\widehat{\theta}^{\mathrm{TG}}_{n,B_n}$ is the total-Gram estimator in Algorithm 1, h_n is the information-balance bandwidth for δ_n in Definition 8, and the rate is the frontier in Definition 9.

Proof of Lemma 15. Let $\ell = \lceil \beta \rceil - 1$. From Lemma 17, choose constants $C_T, c_T > 0$ such that the population Gram constant $\lambda_\star = \lambda_\star(\ell, \kappa, c_-, c_+)$ is positive and the following bound holds uniformly over $P \in \mathcal{M}$, sample sizes, admissible splits, thresholds $\delta \in [0, \bar{\delta}]$, and strata $x \in \mathcal{X}$: if h is the information-balance bandwidth from Definition 8 evaluated at $(n, \delta, \beta, \kappa, \bar{\delta})$, then

$$P^{\otimes n}(\Omega_x^c) \le C_T \exp\{-c_T n h (\delta + h)^\kappa\}.$$

Define

$$K_{\text{noise}} = \frac{1}{2}\left(\frac{336\{4(\ell+1)/\lambda_\star^2\}}{p_{\min} c_- /(2 \cdot 2^\kappa)}\right)^{1/2}, \qquad K_{\text{bias}} = \frac{2\sqrt{\ell+1}}{\lambda_\star},$$
$$K_{\text{loc}} = K_{\text{noise}} + L K_{\text{bias}} + \frac{C_T}{c_T}, \qquad K_{\text{atom}} = \frac{c_+}{\kappa+1},$$

and

$$C_{\text{up}} = 2(1+J)(1 + K_{\text{atom}} K_{\text{loc}}).$$

The displayed constants are finite and nonnegative, and $C_{\text{up}} > 0$, by the regime assumptions and $\lambda_\star > 0$.

Fix (δ_n) satisfying $\delta_n \in [0, \bar{\delta}]$ for every n, and fix an admissible split sequence $B_\bullet$. By Lemma 10, for all sufficiently large n,

$$0 < h_n, \qquad h_n \le 1 - \bar{\delta}, \qquad n h_n^{2\beta+1}(\delta_n + h_n)^\kappa = 1.$$

After increasing N, also $n \ge 8$. Then $\delta_n + h_n \le 1$. Put

$$r_{0,n} = n^{-1/2}, \qquad a_n = \delta_n^{\kappa+1} h_n^\beta.$$

Since $|I_{0,n}|, |I_{1,n}| \ge \lfloor n/4 \rfloor$ and $n \ge 8$, we have

$$n \le 8|I_{0,n}|, \qquad n \le 8|I_{1,n}|,$$

and therefore

$$|I_{0,n}|^{-1/2} \le 3n^{-1/2}, \qquad |I_{1,n}|^{-1/2} \le 3n^{-1/2}.$$

The balance identity gives

$$n h_n (\delta_n + h_n)^\kappa = h_n^{-2\beta}.$$

Thus, using $h_n \le 1$ and $\beta > 0$,

$$C_T \exp\{-c_T n h_n (\delta_n + h_n)^\kappa\} \le \frac{C_T}{c_T} h_n^\beta.$$

Indeed, with $e = n h_n (\delta_n + h_n)^\kappa = h_n^{-2\beta}$, the elementary bound $\exp(-c_T e) \le (c_T e)^{-1}$ gives $(C_T/c_T) h_n^{2\beta}$, and $h_n^{2\beta} \le h_n^\beta$.

For any $P \in \mathcal{M}$, the projection $\Pi_{[0,1]}$ is one-Lipschitz and the finite sum over strata gives the deterministic decomposition

$$\mathbb{E}_{P^{\otimes n}} \left|\widehat{\theta}^{\text{TG}}_{n,B_n} - \theta_{\delta_n}(P)\right| \le \mathbb{E}|\widehat{\nu}_n - \nu_{\delta_n}(P)| + \sum_{x \in \mathcal{X}} \left\{\mathbb{E}|\widehat{b}_x - p_x q_x(\delta_n)| + |p_x q_x(\delta_n)| \mathbb{E}|\widetilde{\mu}_x - \mu_x^P(\delta_n)|\right\},$$

where all expectations on the right are under $P^{\otimes n}$. The retained-course and atom-mass block averages are averages of $[0,1]$-valued variables, hence

$$\mathbb{E}|\widehat{\nu}_n - \nu_{\delta_n}(P)| \le \frac{1}{2}|I_{0,n}|^{-1/2}, \qquad \mathbb{E}|\widehat{b}_x - p_x q_x(\delta_n)| \le \frac{1}{2}|I_{1,n}|^{-1/2}.$$

By Proposition 6 and $0 \le p_x \le 1$,

$$|p_x q_x(\delta_n)| \le \frac{c_+ \delta_n^{\kappa+1}}{\kappa+1} = K_{\mathrm{atom}} \delta_n^{\kappa+1}.$$

It remains to bound the local regression term. On Ω_x, polynomial reproduction and the deterministic weight bound in Lemma 12 give the Hölder bias bound

$$\left| \sum_{i \in I_{2,n}} w_{ix} \mu_X^P(A_i) - \mu_x^P(\delta_n) \right| \le L K_{\mathrm{bias}} h_n^{\beta}.$$

The centered noise term satisfies

$$\mathbb{E} \left| \sum_{i \in I_{2,n}} w_{ix} \{Y_i - \mu_X^P(A_i)\} \right| \le K_{\mathrm{noise}} h_n^{\beta},$$

because the conditional variance is bounded by $\frac{1}{4}\sum_i w_{ix}^2$ and Lemma 13 controls the integrated square-root weight energy. On Ω_x^c, both the fallback value and $\mu_x^P(\delta_n)$ lie in $[0,1]$, so the exceptional contribution is at most $P^{\otimes n}(\Omega_x^c)$. Combining the good-event decomposition with the noise, bias, and tail bounds gives

$$\mathbb{E}|\widetilde{\mu}_x - \mu_x^P(\delta_n)| \le K_{\mathrm{noise}} h_n^{\beta} + L K_{\mathrm{bias}} h_n^{\beta} + C_T \exp\{-c_T n h_n (\delta_n + h_n)^{\kappa}\} \le K_{\mathrm{loc}} h_n^{\beta}.$$

Before the split-size and tail simplifications, the finite-sample bound is

$$\mathbb{E}_{P^{\otimes n}} \left| \widehat{\theta}_{n,B_n}^{\mathrm{TG}} - \theta_{\delta_n}(P) \right| \le \frac{1}{2}|I_{0,n}|^{-1/2} + J \left\{ \frac{1}{2}|I_{1,n}|^{-1/2} + K_{\mathrm{atom}} \delta_n^{\kappa+1} \Big(K_{\mathrm{noise}} h_n^{\beta} + L K_{\mathrm{bias}} h_n^{\beta} + C_T \exp\{-c_T n h_n (\delta_n +$$

Using the two split-block inequalities above and the local bound just obtained, uniformly in $P \in \mathcal{M}$,

$$\mathbb{E}_{P^{\otimes n}} \left| \widehat{\theta}_{n,B_n}^{\mathrm{TG}} - \theta_{\delta_n}(P) \right| \le 2 r_{0,n} + J\{2 r_{0,n} + K_{\mathrm{atom}} K_{\mathrm{loc}} a_n\}.$$

The model class is nonempty. Take the centered Bernoulli law with $p_x = 1/J$, conditional treatment density $\pi_x(a) = (\kappa+1)a^{\kappa}$, and outcome distribution Bernoulli with mean $1/2$ conditional on (X, A); this law has regression $\mu_x^P \equiv 1/2$ and belongs to $\mathcal{M}$ under the regime assumptions. Hence the same upper bound applies to the supremum over $P \in \mathcal{M}$. Since $r_{0,n}, a_n \ge 0$,

$$2 r_{0,n} + J\{2 r_{0,n} + K_{\mathrm{atom}} K_{\mathrm{loc}} a_n\} \le 2(1+J)(1 + K_{\mathrm{atom}} K_{\mathrm{loc}})(r_{0,n} + a_n).$$

Finally, $r_{0,n} + a_n$ is exactly the frontier rate in Definition 9. This proves the asserted bound with the displayed C_{up}. □

Lemma 8 states the monomial reproduction property in the same notation used by the estimator. This property turns local-polynomial weighting into an estimate of the regression value at the threshold.

Lemma 16. [`lem:bandwidth-phases`] (Bandwidth phase bounds) *There exist constants $c_{\mathrm{bal}}, C_{\mathrm{bal}}, c_{\mathrm{far}}, C_{\mathrm{far}}$ with*

$$0 < c_{\mathrm{bal}} \le C_{\mathrm{bal}}, \qquad 0 < c_{\mathrm{far}} \le C_{\mathrm{far}},$$

such that the following assertions hold. Let

$$\delta_{\mathrm{edge},n} = n^{-1/(2\beta+\kappa+1)}.$$

For every J, every observed law P, every $L, c_-, c_+, p_{\min}, \alpha$, and every deterministic sequence $(\delta_n)_{n\ge1}$, suppose that:

- ***(Regime constants.)** $J \ge 1$, $\beta > 0$, $\kappa \ge 0$, $L > 0$, $c_- > 0$, $c_- \le \kappa+1 \le c_+$, $p_{\min} > 0$, $p_{\min} \le 1/J$, $0 < \bar\delta < 1$, and $0 < \alpha < 1/2$.*
- ***(Thinning.)** The law P satisfies the polynomial thinning condition in Assumption 4 with exponent κ and constants c_-, c_+.*
- ***(Thresholds.)** $\delta_n \in [0, \bar\delta]$ for every n.*

Let h_n be the information-balance bandwidth from Definition 8 for $(n, \delta_n, \beta, \kappa, \bar\delta)$. Then, for all sufficiently large n,

$$c_{\mathrm{bal}} \le n h_n^{2\beta+1}(\delta_n + h_n)^\kappa \le C_{\mathrm{bal}},$$

and

$$c_{\mathrm{bal}} h_n^{-2\beta} \le n h_n (\delta_n + h_n)^\kappa \le C_{\mathrm{bal}} h_n^{-2\beta}.$$

If $\delta_n/\delta_{\mathrm{edge},n} \to \infty$, then, for all sufficiently large n,

$$c_{\mathrm{far}}\left(n\delta_n^\kappa\right)^{-1/(2\beta+1)} \le h_n \le C_{\mathrm{far}}\left(n\delta_n^\kappa\right)^{-1/(2\beta+1)}.$$

Moreover, for every $D > 0$, there exist constants $c_{\mathrm{near}}, C_{\mathrm{near}}$, depending on D, with

$$0 < c_{\mathrm{near}} \le C_{\mathrm{near}},$$

such that, under the same regime-constant, thinning, and threshold conditions, if

$$|\delta_n| \le D|\delta_{\mathrm{edge},n}|$$

for all sufficiently large n, then, for all sufficiently large n,

$$c_{\mathrm{near}} n^{-1/(2\beta+\kappa+1)} \le h_n \le C_{\mathrm{near}} n^{-1/(2\beta+\kappa+1)}.$$

Proof of Lemma 16. Set

$$p = 2\beta+1, \qquad c_{\mathrm{bal}} = C_{\mathrm{bal}} = 1, \qquad C_{\mathrm{far}} = 1,$$

and choose

$$c_{\mathrm{far}} = \min\{1, 2^{-\kappa/p}\}.$$

At this initial choice of witnesses, the quotient $-\kappa/p$ is interpreted by the total real-division convention: it equals 0 when $p = 0$ and agrees with ordinary division when $p \ne 0$. Thus c_{far} is a real constant for every value of β. Since the base 2 is positive, $2^{-\kappa/p} > 0$, and therefore

$$0 < c_{\mathrm{far}} \le 1 = C_{\mathrm{far}}.$$

Together with $c_{\text{bal}} = C_{\text{bal}} = 1$, this gives

$$0 < c_{\text{bal}} \le C_{\text{bal}}, \qquad 0 < c_{\text{far}} \le C_{\text{far}}.$$

Fix $J, P, L, c_-, c_+, p_{\min}, \alpha$ and a deterministic threshold sequence satisfying the hypotheses. The regime assumptions give $\beta > 0$, $\kappa \ge 0$, and $0 < \bar{\delta} < 1$, hence $p = 2\beta + 1 > 0$. Let h_n be the information-balance bandwidth from Definition 8; thus

$$h_n = \inf\{h \in (0, 1 - \bar{\delta}] : nh^{2\beta+1}(\delta_n + h)^\kappa \ge 1\}$$

when the displayed set is nonempty, with fallback value $h_n = 1 - \bar{\delta}$ when that set is empty. By Lemma 10, for all sufficiently large n,

$$0 < h_n, \qquad h_n \le 1 - \bar{\delta}, \qquad nh_n^p(\delta_n + h_n)^\kappa = 1.$$

The first balance assertion follows immediately:

$$1 \le nh_n^{2\beta+1}(\delta_n + h_n)^\kappa \le 1.$$

Since $h_n > 0$,

$$h_n^{2\beta+1} = h_n^{2\beta} h_n, \qquad h_n^{-2\beta} = (h_n^{2\beta})^{-1}.$$

Dividing the balance identity by $h_n^{2\beta}$ gives

$$nh_n(\delta_n + h_n)^\kappa = h_n^{-2\beta},$$

and therefore

$$h_n^{-2\beta} \le nh_n(\delta_n + h_n)^\kappa \le h_n^{-2\beta}.$$

Now assume

$$\frac{\delta_n}{\delta_{\text{edge},n}} \to \infty, \qquad \delta_{\text{edge},n} = n^{-1/(2\beta+\kappa+1)} = n^{-1/(p+\kappa)}.$$

On a sufficiently large eventual set, the balance identity holds, $n > 0$, and the convergence gives $1 \le \delta_n/\delta_{\text{edge},n}$. Since $\delta_{\text{edge},n} > 0$ on this set,

$$\delta_{\text{edge},n} \le \delta_n, \qquad \delta_n > 0.$$

For those n, $\delta_n^\kappa \le (\delta_n + h_n)^\kappa$, so the balance identity implies

$$h_n^p(n\delta_n^\kappa) \le 1.$$

Taking p-th roots gives

$$h_n \le (n\delta_n^\kappa)^{-1/p}.$$

The comparison $\delta_{\text{edge},n} \le \delta_n$ gives

$$1 = n\delta_{\text{edge},n}^{p+\kappa} \le n\delta_n^{p+\kappa} = \delta_n^p(n\delta_n^\kappa),$$

and hence

$$(n\delta_n^\kappa)^{-1/p} \le \delta_n.$$

Consequently $h_n \le \delta_n$, and therefore

$$\delta_n + h_n \le 2\delta_n.$$

Using the balance identity again,

$$1 = nh_n^p(\delta_n + h_n)^\kappa \le nh_n^p(2\delta_n)^\kappa = h_n^p\, 2^\kappa(n\delta_n^\kappa).$$

Thus

$$2^{-\kappa/p}(n\delta_n^\kappa)^{-1/p} \le h_n.$$

Under $p > 0$ and $\kappa \ge 0$, $2^{-\kappa/p} \le 1$, so the chosen constant satisfies $c_{\text{far}} = 2^{-\kappa/p}$. With $C_{\text{far}} = 1$, the far-threshold comparison follows:

$$c_{\text{far}}(n\delta_n^\kappa)^{-1/(2\beta+1)} \le h_n \le C_{\text{far}}(n\delta_n^\kappa)^{-1/(2\beta+1)}.$$

Finally fix $D > 0$. Set

$$C_{\text{near}} = 1, \qquad c_{\text{near}} = \min\{1, (D+1)^{-\kappa/p}\},$$

again using the same total real-division convention for the displayed quotient. Since $D + 1 > 0$, the positive-base property of real powers gives $(D+1)^{-\kappa/p} > 0$, and therefore

$$0 < c_{\text{near}} \le 1 = C_{\text{near}}.$$

Fix again $J, P, L, c_-, c_+, p_{\min}, \alpha$ and a deterministic threshold sequence satisfying the hypotheses, and let h_n be the corresponding bandwidth from Definition 8. The regime assumptions again give $p > 0$ and $\kappa \ge 0$. Suppose that eventually

$$|\delta_n| \le D|\delta_{\text{edge},n}|.$$

On the eventual set where this bound and Lemma 10 both hold, $n > 0$, $\delta_n \ge 0$, and $\delta_{\text{edge},n} > 0$, so

$$\delta_n \le Dn^{-1/(p+\kappa)}.$$

The balance identity and $h_n^\kappa \le (\delta_n + h_n)^\kappa$ give

$$nh_n^{p+\kappa} \le 1,$$

hence

$$h_n \le n^{-1/(p+\kappa)}.$$

Together with the threshold bound, this gives

$$\delta_n + h_n \le (D+1)n^{-1/(p+\kappa)}.$$

Therefore

$$1 = nh_n^p(\delta_n + h_n)^\kappa \le nh_n^p(D+1)^\kappa\big(n^{-1/(p+\kappa)}\big)^\kappa.$$

Since $n\big(n^{-1/(p+\kappa)}\big)^{p+\kappa} = 1$, the displayed inequality $1 \le nh_n^p(D+1)^\kappa(n^{-1/(p+\kappa)})^\kappa$ rearranges to

$$(D+1)^{-\kappa/p}n^{-1/(p+\kappa)} \le h_n.$$

Under $p > 0$, $\kappa \ge 0$, and $D > 0$, $(D+1)^{-\kappa/p} \le 1$, so the chosen constant satisfies $c_{\text{near}} = (D+1)^{-\kappa/p}$. Since $p + \kappa = 2\beta + \kappa + 1$, we obtain, for all sufficiently large n,

$$c_{\text{near}}n^{-1/(2\beta+\kappa+1)} \le h_n \le C_{\text{near}}n^{-1/(2\beta+\kappa+1)}.$$

□

Lemma 9 gives the pointwise weight control implied by Gram coercivity. Together with Lemma 12, it bounds both individual leverage and aggregate variance.

Lemma 17. [`lem:total-gram`] (Total Gram stabilization) *Let J be a positive integer and let the declared constants satisfy*

$$\beta > 0, \qquad \kappa \ge 0, \qquad L > 0, \qquad 0 < c_- \le \kappa{+}1 \le c_+, \qquad 0 < p_{\min} \le 1/J, \qquad 0 < \bar{\delta} < 1, \qquad 0 < \alpha < 1/2.$$

Then there are constants $C, c > 0$, depending only on these declared constants, such that the uniform population Gram lower-bound constant $\lambda_\star$ in Definition 12, evaluated at (ℓ, κ, c_-, c_+) with $\ell = \lceil \beta \rceil - 1$, satisfies $\lambda_\star > 0$, and the following hold.

- ***(Model and sampling.)*** *$P \in \mathcal{M}$ as in Definition 5, and $O_1, \ldots, O_n$ are sampled according to Assumption 1.*
- ***(Split and threshold.)*** *$I_0, I_1, I_2 \subset \{1, \ldots, n\}$ are deterministic, pairwise-disjoint split blocks with $|I_b| \ge \lfloor n/4 \rfloor$ for $b = 0, 1, 2$, and $\delta \in [0, \bar{\delta}]$.*
- ***(Bandwidth and stratum.)*** *h_n is the information-balance bandwidth from Definition 8 at $(n, \delta, \beta, \kappa, \bar{\delta})$, and $x \in \mathcal{X}$.*

For the good-design event Ω_x in Definition 12,

$$P^{\otimes n}(\Omega_x^c) \le C \exp\{-c\, n\, h_n (\delta + h_n)^\kappa\}.$$

On Ω_x, the exact local-polynomial intercept weights w_{ix} in Definition 12 satisfy, for every $j = 0, \ldots, \ell$,

$$\sum_{i \in I_2} w_{ix} U_i^j = \begin{cases} 1, & j = 0, \\ 0, & j \ge 1, \end{cases}$$

and also obey

$$\sum_{i \in I_2} |w_{ix}| \le C, \qquad \sum_{i \in I_2} w_{ix}^2 \le \frac{C}{N_x}.$$

Proof of Lemma 17. Set

$$\ell = \lceil \beta \rceil - 1, \qquad \lambda = \lambda_\star(\ell, \kappa, c_-, c_+), \qquad d = \ell + 1.$$

The regime assumptions give $\kappa \ge 0$, $c_- > 0$, and $c_+ > 0$, so $\lambda > 0$ by Lemma 2.

Define

$$\eta = \min\left\{\frac{1}{20}, \frac{\lambda^2}{16d(4d + \lambda)}\right\}, \qquad K = \frac{p_{\min} c_-}{2 \cdot 2^\kappa}, \qquad c = \frac{\eta K}{8},$$

and

$$A_{\text{tail}} = 2(1 + d^2), \qquad C = A_{\text{tail}} e^{8c} + \frac{2\sqrt{d}}{\lambda} + \frac{4d}{\lambda^2} + 1.$$

The constants $\eta, K, c, A_{\text{tail}}, C$ are positive and, together with λ, depend only on the displayed regime constants.

Fix P, n, B_n, δ, x as in the statement, and write

$$h = h_n, \qquad q(O) = \mathbf{1}\left\{X = x,\ 0 \le \frac{A - \delta}{h} \le 1\right\}, \qquad \phi_j(O) = \mathbf{1}\left\{0 \le \frac{A - \delta}{h} \le 1\right\} \left(\frac{A - \delta}{h}\right)^j.$$

The regime assumptions give $\beta > 0$, $\kappa \ge 0$, $\delta \in [0, \bar{\delta}]$, and $0 < \bar{\delta} < 1$. The construction of h in Definition 8 gives $0 < h \le 1 - \bar{\delta}$. Indeed, if the crossing set in Definition 8 is empty, then $h = 1 - \bar{\delta}$. If it is nonempty and z is a member, then $n > 0$, $0 < z \le 1 - \bar{\delta}$, $\delta + z \le 1$, and

$$1 \le n z^{2\beta+1} (\delta + z)^{\kappa} \le n z^{2\beta+1},$$

so $n^{-1/(2\beta+1)} \le z$. Hence the infimum is bounded below by this positive number and above by any member of the set, giving $0 < h \le 1 - \bar{\delta}$. Since $\delta \le \bar{\delta}$, this also gives $\delta + h \le 1$.

Let

$$p = \mathbb{E}_P q(O) = P\left\{ X = x,\ 0 \le \frac{A - \delta}{h} \le 1 \right\}.$$

By Lemmas 3 and 4 and the thinning and stratum-mass assumptions,

$$p = p_x \int_{\delta}^{\delta+h} \pi_x(a)\, da \ge p_{\min} c_- \int_{\delta}^{\delta+h} a^{\kappa}\, da \ge p_{\min} c_- \frac{h}{2} \left(\frac{\delta + h}{2} \right)^{\kappa} = K\, h (\delta + h)^{\kappa}.$$

For every $b \in \mathbb{R}^{\ell+1}$, Lemma 11 gives

$$\lambda\, p\, \|b\|_2^2 \le \sum_{j=0}^{\ell} \sum_{k=0}^{\ell} b_j b_k\, \mathbb{E}_P\{q(O)\phi_j(O)\phi_k(O)\}.$$

Moreover $0 \le q \le 1$ and $|\phi_j| \le 1$ for every j.

Let $m = |I_2|$, and let $i_1, \ldots, i_m$ be the elements of I_2 in increasing order. For a block vector $\omega = (\omega_1, \ldots, \omega_m)$, define

$$N_{\text{blk}}(\omega) = \sum_{r=1}^{m} q(\omega_r), \qquad M_{\text{blk},jk}(\omega) = \sum_{r=1}^{m} q(\omega_r)\phi_j(\omega_r)\phi_k(\omega_r),$$

and put

$$M_{jk} = \mathbb{E}_P\{q(O)\phi_j(O)\phi_k(O)\}.$$

The deterministic ordered-block map

$$T_{I_2}(O_1, \ldots, O_n) = (O_{i_1}, \ldots, O_{i_m})$$

satisfies, for every measurable block event E,

$$P^{\otimes n}\{T_{I_2}(O_1, \ldots, O_n) \in E\} = P^{\otimes m}(E),$$

because the selected coordinates are independent and each has law P. Under this map,

$$N_{\text{blk}}(T_{I_2} z) = \sum_{i \in I_2} q(O_i) = N_x,$$

and, for every j, k,

$$M_{\text{blk},jk}(T_{I_2} z) = \sum_{i \in I_2} q(O_i)\phi_j(O_i)\phi_k(O_i) = (G_x)_{jk},$$

where the last equality is the entrywise definition of the local Gram matrix in Definition 12.

Define the block good event

$$\Gamma_x = \left\{ \omega : N_{\text{blk}}(\omega) > 0 \text{ and } \sum_{j=0}^{\ell} \sum_{k=0}^{\ell} b_j b_k M_{\text{blk},jk}(\omega) \geq \frac{\lambda N_{\text{blk}}(\omega)}{2} \|b\|_2^2 \text{ for every } b \in \mathbb{R}^d \right\}.$$

The preceding identities give the exact event identity

$$\Omega_x = \{z : T_{I_2} z \in \Gamma_x\}.$$

Indeed, q is the indicator of the stratum-window event, ϕ_j is its local monomial coordinate, and $\lambda = \lambda_\star(\ell, \kappa, c_-, c_+)$, so N_{blk} and M_{blk} are exactly the count and quadratic form appearing in Definition 12.

If $n \geq 8$, the split condition gives $m \geq n/8$, hence $m > 0$. The lower bound on p gives $p > 0$. Applying the localized empirical Gram concentration bound to the m independent block coordinates, with the population coercivity above and with $0 \leq q \leq 1$ and $|\phi_j| \leq 1$, gives

$$P^{\otimes m}(\Gamma_x^c) \leq A_{\text{tail}} \exp\{-mp\eta\}.$$

By the ordered-block product identity and $\Omega_x = \{z : T_{I_2} z \in \Gamma_x\}$,

$$P^{\otimes n}(\Omega_x^c) \leq A_{\text{tail}} \exp\{-mp\eta\}.$$

Combining $m \geq n/8$ with the lower bound on p,

$$c\, n\, h(\delta + h)^\kappa \leq m\, p\, \eta.$$

Hence

$$P^{\otimes n}(\Omega_x^c) \leq A_{\text{tail}} \exp\{-mp\eta\} \leq A_{\text{tail}} \exp\{-c\, n\, h(\delta + h)^\kappa\} \leq C \exp\{-c\, n\, h(\delta + h)^\kappa\}.$$

If $n < 8$, then $h \leq 1$, $\delta + h \leq 1$, and $\kappa \geq 0$ imply

$$n\, h(\delta + h)^\kappa \leq 8.$$

Thus

$$1 \leq A_{\text{tail}} e^{8c} \exp\{-c\, n\, h(\delta + h)^\kappa\} \leq C \exp\{-c\, n\, h(\delta + h)^\kappa\},$$

and the same probability bound follows from $P^{\otimes n}(\Omega_x^c) \leq 1$.

On Ω_x, Lemma 12 gives, for every $0 \leq j \leq \ell$,

$$\sum_{i \in I_2} w_{ix} U_i^j = \mathbf{1}\{j = 0\},$$

which is the displayed reproduction identity. The same result also gives

$$\sum_{i \in I_2} |w_{ix}| \leq \frac{2\sqrt{d}}{\lambda} \leq C, \qquad \sum_{i \in I_2} w_{ix}^2 \leq \frac{4d}{\lambda^2 N_x} \leq \frac{C}{N_x}.$$

□

Lemma 17 assembles the local-window identities, shifted coercivity, and concentration into the good-design probability bound used by the stabilized procedures. The exponential bound makes the singular-design fallback negligible on the scales used in Theorems 1 and 2, while the on-event identities preserve the usual local-polynomial intercept behavior.

The last auxiliary lemma for the appendix supports the continuity-only causal and observed targets. It establishes pointwise uniqueness of the continuous conditional-regression extension on the threshold envelope.

Lemma 18. [`lem:continuity-regression-extension-unique`] (Unique continuous regression extension) *Let J be a positive integer, let $\kappa, c_-, c_+, p_{\min}, \bar{\delta} \in \mathbb{R}$, and let $P \in \mathcal{M}_{\text{cont}}$ be a continuity-only observed clamp model with these constants, as in Definition 24. Suppose that*

- ***(Design constants.)*** *The constants satisfy*

$$0 \le \kappa, \qquad 0 < c_-, \qquad c_- \le \kappa + 1 \le c_+, \qquad 0 < p_{\min} \le \frac{1}{J}, \qquad 0 < \bar{\delta} < 1.$$

- ***(Continuous versions.)*** *For $r = 1, 2$, $\mu_r : \mathcal{X} \times \mathbb{R} \to \mathbb{R}$ satisfies that $a \mapsto \mu_r(x, a)$ is continuous on $[0, \bar{\delta}]$ for every $x \in \mathcal{X}$.*

- ***(Regression version property.)*** *For $r = 1, 2$,*

$$\mathbb{E}_P[Y \mid X, A] = \mu_r(X, A) \quad P\text{-almost surely.}$$

Then, for every $x \in \mathcal{X}$,

$$\mu_1(x, a) = \mu_2(x, a) \qquad \text{for every } a \in [0, \bar{\delta}].$$

Proof of Lemma 18. Fix $x \in \mathcal{X}$. The two regression-version assumptions imply

$$\mu_1(X, A) = \mu_2(X, A) \quad P\text{-a.s.}$$

Define functions on the real line by

$$f_r(a) = \begin{cases} \mu_r(x, a), & a \in [0, \bar{\delta}], \\ 0, & a \notin [0, \bar{\delta}], \end{cases} \qquad r = 1, 2.$$

The continuity hypotheses make these functions measurable. Set

$$B = \{a \in \mathbb{R} : f_1(a) \ne f_2(a)\}.$$

Then B is measurable and

$$B \subset [0, \bar{\delta}] \subset [0, 1].$$

The almost-sure equality of the two versions gives

$$P\{X = x,\ A \in B\} = 0.$$

The conditional-density identity in Assumption 2 applies to this measurable subset of $[0, 1]$, so

$$0 = P\{X = x,\ A \in B\} = p_x \int_B \pi_x(a)\, da.$$

The stratum-mass condition in Assumption 3, together with $p_{\min} > 0$, gives $p_x > 0$. Hence

$$\int_B \pi_x(a)\, da = 0.$$

It remains to justify the zero-integral argument on B. The conditional-density measurability in Assumption 2 gives the needed measurability on the restricted set. Since $B \subset [0,1]$, the set B has finite Lebesgue measure. For Lebesgue-a.e. $a \in B$, the nonnegativity in Assumption 2, the upper envelope in Assumption 4, and the design inequalities $0 \le \kappa$ and $0 \le c_+$ give

$$|\pi_x(a)| = \pi_x(a) \le c_+ a^\kappa \le c_+, \qquad 0 \le a \le 1.$$

Thus $a \mapsto \pi_x(a)$ is integrable on B. The nonnegativity part of Assumption 2 and the zero integral therefore yield

$$\pi_x(a) = 0 \quad \text{for Lebesgue-a.e. } a \in B.$$

The point $a = 0$ has Lebesgue measure zero. For Lebesgue-a.e. $a \in B$ with $a \neq 0$, the inclusion $B \subset [0,1]$ gives $a > 0$. The lower envelope in Assumption 4 and $c_- > 0$ then give

$$\pi_x(a) \ge c_- a^\kappa > 0 \quad \text{for Lebesgue-a.e. } a \in B \setminus \{0\}.$$

Combining this positivity with $\pi_x = 0$ Lebesgue-a.e. on B gives

$$\mathrm{Leb}(B) = 0.$$

Consequently $f_1 = f_2$ Lebesgue-a.e. on $\mathbb{R}$, and the definitions of f_1 and f_2 give

$$\mu_1(x,a) = \mu_2(x,a) \quad \text{for Lebesgue-a.e. } a \in [0, \bar\delta].$$

Finally, both functions $a \mapsto \mu_1(x,a)$ and $a \mapsto \mu_2(x,a)$ are continuous on $[0,\bar\delta]$. If they differed at some point of $[0,\bar\delta]$, continuity would make them differ on a nonempty relatively open interval in $[0,\bar\delta]$, which has positive Lebesgue measure, contradicting the almost-everywhere equality on $[0,\bar\delta]$. Hence

$$\mu_1(x,a) = \mu_2(x,a) \qquad \text{for every } a \in [0,\bar\delta].$$

Since x was arbitrary, the conclusion holds in every stratum. □

Lemma 18 justifies the pointwise notation for the continuity-only regression extension used in Definition 25. The positive treatment-density envelope gives enough support throughout $[0,\bar\delta]$ for almost-everywhere equality of continuous versions to determine the extension at every threshold value.

Verification note. The Lean kernel checks were run at repository commit `6f1bf59b1255336e29c22bdea39edd5` using the CausalSmith Lean toolchain `leanprover/lean4:v4.33.0`; the executable reports Lean 4.33.0, target `x86_64-unknown-linux-gnu`, Lean commit `d8b18978322de05a8f3dba51ef03cf5461676c17`. The kernel-checked assumption, class, definition, and algorithm anchors in this manuscript are the objects introduced earlier in the paper. The kernel-checked result and open-question anchors are the objects introduced earlier in the paper. Presentation-synthesized restatements such as Definitions 1 to 4, 10, 11, 13 to 19, 23, 29, 31 and 32 are reader-facing anchors whose status is recorded separately in the verification contract. Remark 1 contributes a kernel-checked, well-formed encoding of the open sharp-constant question in the formal layer. Statement- and proof-faithfulness audits compare the displayed manuscript statements and proof narratives with those checked declarations; those audits are separate from Lean kernel checking. The theorem-local verification footnotes identify the published conditional-distribution input from Kallenberg [2002] and the bounded-difference concentration input from Hoeffding [1963] at the exact statements where those cited conclusions are invoked.

Reproducing the verification. The material needed to reproduce the checks is organized as follows.

Archived source. The Lean development is released in the public CausalSmith repository at `https://github.com/Jiyuan-Tan/CausalSmith`; the state checked for this manuscript is the commit recorded above, and a snapshot of that commit is deposited in a persistent archive with a DOI at publication so that the checked state remains citable independently of the repository history. All declarations for this paper live in the single module tree `CausalSmith/Stat/STAT_LmtpThresholdAtomFrontier_Research`.

Dependencies. The build needs the Lean toolchain `leanprover/lean4:v4.33.0`, pinned by the repository's `lean-toolchain` file and installed by `elan`, together with the Lean build tool `lake`, which is distributed with the toolchain. The only non-Lean prerequisites are `git` and `curl`. There are two Lean package dependencies, both pinned in the repository's lockfile: the Causalean library, on which this development builds, and Mathlib, on which Causalean builds. No proprietary component, no external solver, and no network access beyond the initial fetch is required.

Build and check. From a clean environment, cloning the repository, checking out the recorded commit, fetching the pinned dependencies with `lake exe cache get`, and building the module tree above with `lake build` reproduces the kernel checks. A successful build is not by itself the claim made here: the reported status additionally requires that the sources contain no `sorry`, no `admit`, and no added `axiom`, and that `#print axioms` on each anchored declaration reports only the three standard Lean axioms. Both checks are scripted in the repository and are rerun against the archived commit.

Anchor-to-declaration manifest. The bundle accompanying this manuscript ships a machine-readable crosswalk that maps every numbered environment in the paper to the declaration that certifies it. Each record carries the manuscript anchor, the environment type and displayed number, the fully qualified Lean name, the declarations it uses, and a status field separating kernel-checked anchors from the presentation-synthesized restatements listed above, which carry no standalone declaration. A companion file records, for each anchor, the verbatim source text of the certifying declaration, so a reader can compare a displayed statement with the checked one without rebuilding the library. The interactive edition of this paper renders that manifest inline at each numbered environment.

B Proofs of the main results

Proof of Theorem 1. Fix the displayed regime constants. We use seven steps.

1. Write $R_n^\star$ for the observed minimax risk over the admissible estimator class used here,

$$R_n^\star = \inf_{T_n} \sup_{P \in \mathcal{M}} \mathbb{E}_{P^{\otimes n}} \left[|T_n - \theta_{\delta_n}(P)| \right],$$

where the infimum ranges over observed-sample measurable estimators satisfying $T_n(z) \in [0,1]$ for every observed sample z. For a split B_n, put

$$\mathcal{R}_n^{\mathrm{TG}}(B_n) = \sup_{P \in \mathcal{M}} \mathbb{E}_{P^{\otimes n}} \left[\left| \widehat{\theta}_{n,B_n}^{\mathrm{TG}} - \theta_{\delta_n}(P) \right| \right].$$

This is the finite-sample stabilized risk in Definition 16, with the product sample law interpreted as in Definition 2. Lemma 15 gives a constant $C_{\mathrm{up}} > 0$ such that, for every admissible deterministic (δ_n) and $B_\bullet$, for all sufficiently large n,

$$\mathcal{R}_n^{\mathrm{TG}}(B_n) \le C_{\mathrm{up}} \left(n^{-1/2} + \delta_n^{\kappa+1} h_n^{\beta} \right).$$

The estimator $\widehat{\theta}_{n,B_n}^{\mathrm{TG}}$ is observed-sample measurable and $[0,1]$-valued. Hence it is an admissible competitor in the infimum defining $R_n^\star$, so

$$R_n^\star \le \mathcal{R}_n^{\mathrm{TG}}(B_n).$$

2. We record the two-point device used for both lower bounds. Let $H_0, H_1 \in \mathcal{M}$, put $\vartheta_j = \theta_\delta(H_j)$, and suppose

$$0 \le \Delta \le |\vartheta_1 - \vartheta_0|, \qquad 0 \le \delta \le 1,$$

with well-posed product chi-squared divergence

$$H_1^{\otimes n} \ll H_0^{\otimes n}, \qquad \int \left(\frac{dH_1^{\otimes n}}{dH_0^{\otimes n}} - 1 \right)^2 dH_0^{\otimes n} < \infty,$$

and

$$\chi^2(H_1^{\otimes n}, H_0^{\otimes n}) \le \chi, \qquad 0 \le \chi < 4.$$

For any admissible $[0,1]$-valued observed-sample estimator T_n, the testing inequality and $\|H_1^{\otimes n} - H_0^{\otimes n}\|_{\mathrm{TV}} \le \frac{1}{2}\sqrt{\chi}$ give

$$\max_{j=0,1} H_j^{\otimes n} \left\{ |T_n - \vartheta_j| \ge \Delta/2 \right\} \ge \frac{1 - \frac{1}{2}\sqrt{\chi}}{2}.$$

Multiplying by $\Delta/2$, the larger of the two risks under H_0, H_1 is at least

$$\frac{1 - \frac{1}{2}\sqrt{\chi}}{4} \Delta.$$

The supremum over $P \in \mathcal{M}$ in the restricted minimax problem dominates these two risks for each T_n; taking the infimum over admissible T_n yields

$$R_n^\star \ge \frac{1 - \frac{1}{2}\sqrt{\chi}}{4} \Delta.$$

3. For the parametric component, use the canonical design

$$P(X = x) = 1/J, \qquad A \mid X = x \sim (\kappa+1) a^\kappa \, da, \quad a \in [0,1],$$

and Bernoulli regressions

$$\mu_x^{P_{\mathrm{r},0}}(a) = \frac{1}{2}, \qquad \mu_x^{P_{\mathrm{r},1}}(a) = \frac{1}{2} + \varepsilon_n, \qquad \varepsilon_n = \frac{1}{4} n^{-1/2}.$$

Both laws belong to $\mathcal{M}$. The stratum masses are $1/J$, which meets the stratum-mass condition of Assumption 3 under the regime constraint $p_{\min} \le 1/J$; the canonical density $(\kappa+1)a^{\kappa}$ meets the conditional-density condition of Assumption 2 and, by the regime constraint $c_- \le \kappa + 1 \le c_+$, the polynomial-thinning envelope of Assumption 4; and each regression is constant, so every Taylor remainder in Assumption 5 vanishes and the Hölder condition holds for any $L > 0$. Finally $|\varepsilon_n| \le 1/4$ for all $n \ge 1$, so both constant regressions take values in $[0, 1]$. The clamp target shifts by

$$|\theta_{\delta_n}(P_{\mathrm{r},1}) - \theta_{\delta_n}(P_{\mathrm{r},0})| = \varepsilon_n.$$

The product chi-squared divergence is well posed, and

$$1 + \chi^2(P_{\mathrm{r},1}^{\otimes n}, P_{\mathrm{r},0}^{\otimes n}) = (1 + 4\varepsilon_n^2)^n \le \exp(n\, 4\varepsilon_n^2) = \exp(1/4).$$

With $\chi_{\mathrm{r}} = \exp(1/4) - 1 < 4$, the two-point mixture bound $(1 + 4\varepsilon_n^2)^n \le \exp(1/4)$ gives a constant $c_{\mathrm{r}} > 0$, depending only on the regime constants, such that

$$c_{\mathrm{r}} n^{-1/2} \le R_n^{\star}$$

for every admissible (δ_n) and all sufficiently large n.

4. For the localized lower-bound component, use a fixed C^{∞} dose bump $b_{\circ} : \mathbb{R} \to [0, 1]$ centered at zero, with

$$b_{\circ}(0) = 1, \qquad 0 \le b_{\circ}(u) \le 1, \qquad b_{\circ}(u) = 0 \quad \text{for } u \notin [-1, 1].$$

Let $\ell = \ell(\beta)$. Since $\beta > 0$, the number $\gamma_{\beta} = \beta - \ell$ belongs to $(0, 1]$. Because $b_{\circ}$ is smooth and compactly supported, there is a finite constant $K_{\mathrm{bump}} \ge 1$ such that

$$\left| b_{\circ}^{(\ell)}(u) - b_{\circ}^{(\ell)}(v) \right| \le K_{\mathrm{bump}} |u - v|^{\gamma_{\beta}}, \qquad u, v \in \mathbb{R}.$$

Choose $a_{\mathrm{low}} > 0$ so small that

$$a_{\mathrm{low}} \le \frac{1}{4}, \qquad a_{\mathrm{low}} K_{\mathrm{bump}} \le L.$$

For any $0 \le \delta \le 1$ and $0 < h \le 1$, define

$$q_{\delta,h}^{\mathrm{low}}(a) = a_{\mathrm{low}} h^{\beta} b_{\circ}\left(\frac{a - \delta}{h}\right), \qquad f_{\delta,h}^{\mathrm{low}}(a) = \frac{1}{2} + q_{\delta,h}^{\mathrm{low}}(a).$$

Then $f_{\delta,h}^{\mathrm{low}}$ is continuous on $[0, 1]$, and $f_{\delta,h}^{\mathrm{low}}(a) \in [0, 1]$ there because

$$0 \le q_{\delta,h}^{\mathrm{low}}(a) \le a_{\mathrm{low}} h^{\beta} \le a_{\mathrm{low}} \le \frac{1}{4}.$$

For $s, t \in [0, 1]$, differentiating the scaled bump ℓ times and using $\gamma_{\beta} = \beta - \ell$ gives

$$\begin{aligned}\left| \frac{d^{\ell}}{da^{\ell}} q_{\delta,h}^{\mathrm{low}}(s) - \frac{d^{\ell}}{da^{\ell}} q_{\delta,h}^{\mathrm{low}}(t) \right| &\le a_{\mathrm{low}} h^{\beta-\ell} K_{\mathrm{bump}} \left| \frac{s - \delta}{h} - \frac{t - \delta}{h} \right|^{\gamma_{\beta}} \\ &= a_{\mathrm{low}} K_{\mathrm{bump}} |s - t|^{\beta-\ell} \le L |s - t|^{\beta-\ell}.\end{aligned}$$

The one-dimensional Taylor remainder theorem on $[0,1]$ therefore yields

$$\left| f_{\delta,h}^{\text{low}}(t) - \sum_{j=0}^{\ell} \frac{(f_{\delta,h}^{\text{low}})^{(j)}(s)}{j!}(t-s)^j \right| \le L|t-s|^{\beta}, \qquad s,t \in [0,1].$$

Thus the localized Bernoulli regression satisfies the member condition in Assumption 5 uniformly over $0 \le \delta \le 1$ and $0 < h \le 1$. Set

$$\zeta_{\text{low}} = 8(\kappa+1)a_{\text{low}}^2, \qquad \rho_\beta = 2\beta+1, \qquad \tau_{\text{low}} = \exp\left(-\frac{\zeta_{\text{low}}+1}{\rho_\beta}\right).$$

Then $0 < \tau_{\text{low}} \le 1$ and

$$\zeta_{\text{low}}\tau_{\text{low}}^{2\beta+1} \le 1.$$

For the contracted bandwidth $\tau_{\text{low}}h_n$, form Bernoulli laws

$$\mu_x^{Q_{\text{low},0}}(a) = \frac{1}{2}, \qquad \mu_x^{Q_{\text{low},1}}(a) = \frac{1}{2} + a_{\text{low}}(\tau_{\text{low}}h_n)^{\beta} b_\circ\left(\frac{a-\delta_n}{\tau_{\text{low}}h_n}\right)$$

on the same canonical design. For all sufficiently large n, Lemma 10 gives

$$0 < h_n, \qquad h_n \le 1-\bar{\delta}, \qquad nh_n^{2\beta+1}(\delta_n+h_n)^{\kappa} = 1.$$

Together with $\delta_n \in [0,\bar{\delta}]$, this gives $\delta_n \in [0,1]$, $0 < \tau_{\text{low}}h_n \le h_n \le 1$, and $\delta_n + h_n \le 1$. The displayed Taylor bound places both laws in $\mathcal{M}$. The target separation obeys

$$a_{\text{low}}\delta_n^{\kappa+1}(\tau_{\text{low}}h_n)^{\beta} \le \left|\theta_{\delta_n}(Q_{\text{low},1}) - \theta_{\delta_n}(Q_{\text{low},0})\right|.$$

The one-observation chi-squared increment satisfies

$$\int 4a_{\text{low}}^2(\tau_{\text{low}}h_n)^{2\beta} b_\circ\left(\frac{a-\delta_n}{\tau_{\text{low}}h_n}\right)^2 dP_{X,A}(x,a) \le 8(\kappa+1)a_{\text{low}}^2(\tau_{\text{low}}h_n)^{2\beta+1}(\delta_n+\tau_{\text{low}}h_n)^{\kappa}.$$

Using the balance identity above, $\tau_{\text{low}}h_n \le h_n$, and $\kappa \ge 0$,

$$n\int 4a_{\text{low}}^2(\tau_{\text{low}}h_n)^{2\beta} b_\circ\left(\frac{a-\delta_n}{\tau_{\text{low}}h_n}\right)^2 dP_{X,A}(x,a) \le \zeta_{\text{low}}\tau_{\text{low}}^{2\beta+1} \le 1.$$

Let

$$I_{\text{low},n} = \int 4a_{\text{low}}^2(\tau_{\text{low}}h_n)^{2\beta} b_\circ\left(\frac{a-\delta_n}{\tau_{\text{low}}h_n}\right)^2 dP_{X,A}(x,a)$$

be the one-observation chi-squared increment for this Bernoulli perturbation. The product law factorization gives

$$1 + \chi^2(Q_{\text{low},1}^{\otimes n}, Q_{\text{low},0}^{\otimes n}) = (1+I_{\text{low},n})^n.$$

Since $I_{\text{low},n} \ge 0$ and $nI_{\text{low},n} \le 1$, the inequality $1+u \le e^u$ yields

$$\chi^2(Q_{\text{low},1}^{\otimes n}, Q_{\text{low},0}^{\otimes n}) \le \exp(nI_{\text{low},n}) - 1 \le e-1 < 4.$$

Applying the two-point bound gives a constant $c_{\text{loc}} > 0$ such that

$$c_{\text{loc}}\delta_n^{\kappa+1}h_n^{\beta} \le R_n^\star$$

for every admissible (δ_n) and all sufficiently large n.

5. The global witness in the theorem uses the same canonical design and

$$\mu_x^{P_0}(a) = \frac{1}{2}, \qquad \mu_x^{P_1}(a) = \frac{1}{2} + n^{-1/2}.$$

For all sufficiently large n, $n^{-1/2} \le 1/2$, so these are Bernoulli laws in $\mathcal{M}$, with shared (X, A)-design, shared p_x, and shared π_x. They satisfy

$$\mu_x^{P_1}(a) = \mu_x^{P_0}(a) + n^{-1/2}$$

for every x and $a \in [0, 1]$, and

$$\frac{1}{2} n^{-1/2} \le |\theta_{\delta_n}(P_1) - \theta_{\delta_n}(P_0)|\,.$$

The one-observation chi-squared increment is $4n^{-1}$. Hence the product divergence is well posed and satisfies

$$1 + \chi^2(P_1^{\otimes n}, P_0^{\otimes n}) = (1 + 4n^{-1})^n \le \exp(4),$$

so

$$\chi^2(P_1^{\otimes n}, P_0^{\otimes n}) \le \exp(4).$$

6. For the localized witness appearing in the statement, use the same fixed bump $b_\circ$. The amplitude construction above also permits a separate witness amplitude: choose $\eta_{\mathrm{wit}} > 0$ such that

$$\eta_{\mathrm{wit}} \le \frac{1}{4}, \qquad \eta_{\mathrm{wit}} K_{\mathrm{bump}} \le L.$$

For every $0 \le \delta \le 1$ and $0 < h \le 1$, the function

$$f_{\delta,h}^{\mathrm{wit}}(t) = \frac{1}{2} + \eta_{\mathrm{wit}} h^\beta b_\circ\left(\frac{t-\delta}{h}\right)$$

is continuous on $[0, 1]$, takes values in $[0, 1]$, and the same scaled-derivative calculation gives

$$\left| f_{\delta,h}^{\mathrm{wit}}(t) - \sum_{j=0}^{\ell} \frac{(f_{\delta,h}^{\mathrm{wit}})^{(j)}(s)}{j!}(t-s)^j \right| \le L|t-s|^\beta, \qquad s, t \in [0, 1].$$

Set

$$\zeta_{\mathrm{wit}} = 8(\kappa + 1)\eta_{\mathrm{wit}}^2, \qquad D_{\mathrm{wit}} = \exp(\zeta_{\mathrm{wit}}).$$

For h_n itself, define

$$\mu_x^{Q_0}(t) = \frac{1}{2}, \qquad \mu_x^{Q_1}(t) = \frac{1}{2} + \eta_{\mathrm{wit}} h_n^\beta b_\circ\left(\frac{t-\delta_n}{h_n}\right).$$

For all sufficiently large n, Lemma 10 gives

$$0 < h_n, \qquad h_n \le 1 - \bar{\delta}, \qquad n h_n^{2\beta+1}(\delta_n + h_n)^\kappa = 1.$$

Together with $\delta_n \in [0, \bar{\delta}]$, this gives $\delta_n \in [0, 1]$, $h_n \le 1$, and $\delta_n + h_n \le 1$, so $Q_0, Q_1 \in \mathcal{M}$. The pair has Bernoulli outcomes, common design, common p_x, common π_x, and

$$\mu_x^{Q_1}(t) - \mu_x^{Q_0}(t) = \eta_{\mathrm{wit}} h_n^\beta b_\circ\left(\frac{t-\delta_n}{h_n}\right)$$

for every x and $t \in [0,1]$. Since $b_\circ(0) = 1$ and the perturbation is nonnegative,

$$\eta_{\mathrm{wit}} \delta_n^{\kappa+1} h_n^\beta \le |\theta_{\delta_n}(Q_1) - \theta_{\delta_n}(Q_0)| .$$

The same integral calculation and the balance identity give

$$n \int 4\eta_{\mathrm{wit}}^2 h_n^{2\beta} b_\circ \left(\frac{t-\delta_n}{h_n} \right)^2 \, dP_{X,A}(x,t) \le \zeta_{\mathrm{wit}}.$$

Let

$$I_{\mathrm{wit},n} = \int 4\eta_{\mathrm{wit}}^2 h_n^{2\beta} b_\circ \left(\frac{t-\delta_n}{h_n} \right)^2 \, dP_{X,A}(x,t)$$

be the one-observation chi-squared increment. The product law factorization gives

$$1 + \chi^2(Q_1^{\otimes n}, Q_0^{\otimes n}) = (1 + I_{\mathrm{wit},n})^n.$$

The perturbation is bounded by $1/4$, so the Bernoulli likelihood ratio is finite and square-integrable for one observation; absolute continuity and square integrability tensorize to the product laws, giving a well-posed product chi-squared divergence. Since $I_{\mathrm{wit},n} \ge 0$ and $nI_{\mathrm{wit},n} \le \zeta_{\mathrm{wit}}$, the inequality $1 + u \le e^u$ gives

$$\chi^2(Q_1^{\otimes n}, Q_0^{\otimes n}) \le \exp(nI_{\mathrm{wit},n}) - 1 \le \exp(\zeta_{\mathrm{wit}}) - 1 \le D_{\mathrm{wit}}.$$

7. Now set

$$c = \frac{1}{2} \min\{c_{\mathrm{r}}, c_{\mathrm{loc}}, 1/2, \eta_{\mathrm{wit}}\}, \qquad C = C_{\mathrm{up}} + D_{\mathrm{wit}} + \exp(4) + c + 1.$$

The amplitude asserted in the theorem is η_{wit}. Then $0 < c < C$, and $c \le 1/2$, $c \le \eta_{\mathrm{wit}}$, $2c \le c_{\mathrm{r}}$, and $2c \le c_{\mathrm{loc}}$. Increasing the eventual index so that all preceding bounds hold, put

$$r_{0,n} = n^{-1/2}, \qquad r_{1,n} = \delta_n^{\kappa+1} h_n^\beta, \qquad r_n = r_{0,n} + r_{1,n}.$$

The two lower components imply

$$R_n^\star \ge c_{\mathrm{r}} r_{0,n}, \qquad R_n^\star \ge c_{\mathrm{loc}} r_{1,n}.$$

Since $r_{0,n}, r_{1,n} \ge 0$, these inequalities and the choices of c give

$$R_n^\star \ge \max\{2c\, r_{0,n}, 2c\, r_{1,n}\} \ge \frac{1}{2}(2c\, r_{0,n}) + \frac{1}{2}(2c\, r_{1,n}) = cr_n.$$

Together with the admissibility and upper bounds above,

$$cr_n \le R_n^\star \le \mathcal{R}_n^{\mathrm{TG}}(B_n) \le Cr_n.$$

By Definitions 2 and 16, $\mathcal{R}_n^{\mathrm{TG}}(B_n)$ is the finite-sample worst-case risk $\mathrm{Risk}_n(B_n)$ of the stabilized estimator computed under the observed i.i.d. product law. The global witness constructed above satisfies

$$cn^{-1/2} \le |\theta_{\delta_n}(P_1) - \theta_{\delta_n}(P_0)| , \qquad \chi^2(P_1^{\otimes n}, P_0^{\otimes n}) \le C,$$

because $c \le 1/2$ and $\exp(4) \le C$. The localized witness constructed above satisfies

$$c\delta_n^{\kappa+1} h_n^\beta \le |\theta_{\delta_n}(Q_1) - \theta_{\delta_n}(Q_0)| , \qquad \chi^2(Q_1^{\otimes n}, Q_0^{\otimes n}) \le C,$$

because $c \le \eta_{\mathrm{wit}}$ and $D_{\mathrm{wit}} \le C$. These are the asserted eventual bounds and witness properties.

□

Proof of Theorem 2. 1. Let $\ell = \lceil \beta \rceil - 1$ and

$$t_\alpha = \sqrt{\frac{\log(12J/\alpha)}{2}}.$$

The expected-length upper bound used below is the following eventual statement: there is a constant $C_{\mathrm{up}} > 0$, depending only on the declared regime constants, such that for every admissible threshold path and split path,

$$\sup_{P \in \mathcal{M}} \mathbb{E}_{P^{\otimes n}}[\mathrm{len}(\mathrm{CI}_n)] \le C_{\mathrm{up}} \left(n^{-1/2} + \delta_n^{\kappa+1} h_n^{\beta} \right)$$

for all sufficiently large n. To see this, use Algorithm 2. Intersecting the symmetric interval with $[0,1]$ weakly shortens its length, and therefore

$$\mathbb{E}_{P^{\otimes n}}[\mathrm{len}(\mathrm{CI}_n)] \le 2\, \mathbb{E}_{P^{\otimes n}} \left[b_{0,n} + \sum_{x \in \mathcal{X}} D_{n,x} \right],$$

where

$$b_{0,n} = \frac{t_\alpha}{\sqrt{|I_{0,n}|}}, \qquad b_{1,n} = \frac{t_\alpha}{\sqrt{|I_{1,n}|}},$$

and $D_{n,x}$ is the stratum contribution in Algorithm 2. The split lower bounds in Definition 1 give $b_{0,n}, b_{1,n} \le 3 t_\alpha n^{-1/2}$ once $n \ge 8$.

For a fixed law $P \in \mathcal{M}$ and stratum x, the integrated stratum contribution is bounded by

$$\mathbb{E}_{P^{\otimes n}} D_{n,x} \le \left(\mathbb{E}_{P^{\otimes n}} \widehat{b}_x + b_{1,n} \right) \left[C_{\ell,1} L h_n^{\beta} + t_\alpha \int \left(\sum_{i \in I_{2,n}} w_{ix}(z)^2 \right)^{1/2} dP^{\otimes n}(z) \right] + b_{1,n} + P^{\otimes n}(\Omega_x^c).$$

Here $C_{\ell,1}$ is the finite ℓ_1-weight constant supplied on Ω_x by Lemma 17. The displayed product bound is the sample-splitting step: the empirical atom estimate uses $I_{1,n}$, the local weights use $I_{2,n}$, and the mixed term factors as

$$\int \widehat{b}_x(z) \left(\sum_{i \in I_{2,n}} w_{ix}(z)^2 \right)^{1/2} dP^{\otimes n}(z) = \left(\int \widehat{b}_x \, dP^{\otimes n} \right) \left(\int \left(\sum_{i \in I_{2,n}} w_{ix}^2 \right)^{1/2} dP^{\otimes n} \right).$$

For the empirical atom envelope, define

$$f_x(O) = \mathbf{1}\{X = x,\ A \le \delta_n\}, \qquad m_x(P) = p_x q_x(\delta_n).$$

Then $\widehat{b}_x = |I_{1,n}|^{-1} \sum_{i \in I_{1,n}} f_x(O_i)$. The mean of one summand is

$$\mathbb{E}_P f_x(O) = P\{X = x, A \le \delta_n\} = p_x \int_0^{\delta_n} \pi_x(a)\, da = p_x q_x(\delta_n) = m_x(P).$$

The second equality is the conditional-density component of Definition 5, and the third is the definition of q_x in Definition 7. Since $0 \le f_x \le 1$, $\mathrm{Var}_P\{f_x(O)\} \le 1/4$. Independence on the deterministic second block therefore yields

$$\begin{aligned}\mathbb{E}_{P^{\otimes n}}\left|\widehat{b}_x - m_x(P)\right| &\le \left\{\mathrm{Var}_{P^{\otimes n}}(\widehat{b}_x)\right\}^{1/2}\\ &\le \frac{1}{2}|I_{1,n}|^{-1/2}.\end{aligned}$$

Moreover, Proposition 6 gives $q_x(\delta_n) \le c_+\delta_n^{\kappa+1}/(\kappa+1)$, and $p_x \le 1$, so

$$m_x(P) \le \frac{c_+}{\kappa+1}\delta_n^{\kappa+1}.$$

The triangle inequality gives the integrated envelope

$$\mathbb{E}_{P^{\otimes n}}\widehat{b}_x \le \frac{1}{2}|I_{1,n}|^{-1/2} + \frac{c_+}{\kappa+1}\delta_n^{\kappa+1}.$$

Also, Lemma 13 gives

$$\int \left(\sum_{i\in I_{2,n}} w_{ix}(z)^2\right)^{1/2} dP^{\otimes n}(z) \le C_{\mathrm{eng}} h_n^{\beta}$$

for a finite constant C_{eng}, and Lemma 17 gives

$$P^{\otimes n}(\Omega_x^c) \le C_{\mathrm{tg}}\exp\{-c_{\mathrm{tg}} n h_n(\delta_n + h_n)^{\kappa}\}.$$

By Lemma 10, for all sufficiently large n,

$$n h_n^{2\beta+1}(\delta_n + h_n)^{\kappa} = 1, \qquad 0 < h_n, \qquad h_n \le 1.$$

The effective local sample size has the following lower bound. Put

$$\gamma_{\mathrm{bal}} = 2\beta + \kappa + 1.$$

Since $\delta_n \ge 0$, $h_n > 0$, and $\kappa \ge 0$,

$$h_n^{\gamma_{\mathrm{bal}}} = h_n^{2\beta+1}h_n^{\kappa} \le h_n^{2\beta+1}(\delta_n + h_n)^{\kappa} = n^{-1}.$$

Because $\gamma_{\mathrm{bal}} > 0$, this gives $h_n \le n^{-1/\gamma_{\mathrm{bal}}}$. Multiplying the balance identity by $h_n^{-2\beta}$ gives

$$n h_n(\delta_n + h_n)^{\kappa} = h_n^{-2\beta},$$

and therefore, since $-2\beta < 0$,

$$n^{2\beta/(2\beta+\kappa+1)} = \left(n^{-1/\gamma_{\mathrm{bal}}}\right)^{-2\beta} \le h_n^{-2\beta} = n h_n(\delta_n + h_n)^{\kappa}.$$

Exponential decay along this polynomial scale is eventually bounded by $n^{-1/2}$. Combining the exponential-decay bound with the polynomial scale bound for $nh_n(\delta_n+h_n)^{\kappa}$, and using $h_n^{\beta} \le 1$, yields uniformly in $P \in \mathcal{M}$ and $x \in \mathcal{X}$,

$$\mathbb{E}_{P^{\otimes n}} D_{n,x} \le C_{\mathrm{rad}}\left(n^{-1/2} + \delta_n^{\kappa+1}h_n^{\beta}\right).$$

Since J is fixed, summing over strata gives the displayed C_{up} bound.

2. Put $\eta = 1 - 2\alpha > 0$. The lower-bound input is a two-point inequality for expected length. Because an arbitrary interval procedure need not have finite expected length, the worst-case length in this step is taken in $[0, \infty]$, with $\int^{-}$ denoting the lower integral of a nonnegative measurable function; for the bounded intervals of interest this is the ordinary expected length. If $P_0, P_1 \in \mathcal{M}$ satisfy

$$\Delta \le |\theta_{\delta_n}(P_1) - \theta_{\delta_n}(P_0)|, \qquad P_1^{\otimes n} \ll P_0^{\otimes n},$$

the likelihood-ratio square is $P_0^{\otimes n}$-integrable, and $\chi^2(P_1^{\otimes n}, P_0^{\otimes n}) \le \chi$, then every uniformly honest observed-sample interval procedure C_n satisfies

$$\max\left\{\left(1 - 2\alpha - \frac{1}{2}\sqrt{\chi}\right)\Delta, 0\right\} \le \sup\left\{\int^{-} \operatorname{len}(C_n)\, dP^{\otimes n} : P \in \mathcal{M}\right\},$$

where the supremum and the lower integral are taken in $[0, \infty]$ and the nonnegative real number on the left is read as an element of $[0, \infty]$ as in Definition 23. The comparisons in this step are made in $[0, \infty]$; the final paragraph returns to the ordinary expected-length notation of Definitions 15 and 18 once the interval is known to be bounded. Indeed, coverage at P_0 and P_1, transferred from $P_1^{\otimes n}$ to $P_0^{\otimes n}$ by

$$d_{\mathrm{TV}}(P_1^{\otimes n}, P_0^{\otimes n}) \le \frac{1}{2}\sqrt{\chi},$$

gives $P_0^{\otimes n}\{\theta_{\delta_n}(P_0), \theta_{\delta_n}(P_1) \in C_n\} \ge 1 - 2\alpha - \frac{1}{2}\sqrt{\chi}$. On that event the interval length is at least Δ, and integrating this indicator gives the displayed extended-real bound.

For the root component, set

$$\varepsilon_n = \tau_\alpha n^{-1/2}, \qquad \tau_\alpha = \frac{\eta}{16}, \qquad \chi_\alpha = \frac{\eta^2}{16}.$$

Let $\mathbb{P}_{0,n}^{\mathrm{root}}$ and $\mathbb{P}_{1,n}^{\mathrm{root}}$ have the common design

$$\mathbb{P}(X = x, A \in da) = \frac{1}{J}(\kappa + 1)a^\kappa \mathbf{1}\{0 \le a \le 1\}\, da,$$

and, conditionally on $(X, A) = (x, a)$, let Y be Bernoulli with success probability

$$\frac{1}{2} + g_{j,n}^{\mathrm{root}}(x, a), \qquad g_{0,n}^{\mathrm{root}}(x, a) = 0, \qquad g_{1,n}^{\mathrm{root}}(x, a) = \varepsilon_n.$$

For all sufficiently large n, $|\varepsilon_n| \le 1/2$. The two laws therefore have outcomes in $[0, 1]$, the same density $\pi_x(a) = (\kappa + 1)a^\kappa$, the same masses $p_x = 1/J$, and constant Hölder regressions; the regime inequalities place both laws in $\mathcal{M}$. Their clamp-target gap is exactly

$$\begin{aligned}\theta_{\delta_n}(\mathbb{P}_{1,n}^{\mathrm{root}}) - \theta_{\delta_n}(\mathbb{P}_{0,n}^{\mathrm{root}}) &= \varepsilon_n \left[\mathbb{P}(A > \delta_n) + \sum_{x \in \mathcal{X}} p_x q_x(\delta_n)\right] \\ &= \varepsilon_n.\end{aligned}$$

For one observation the chi-square divergence is $4\varepsilon_n^2$, hence independence gives

$$1 + \chi^2\big((\mathbb{P}_{1,n}^{\mathrm{root}})^{\otimes n}, (\mathbb{P}_{0,n}^{\mathrm{root}})^{\otimes n}\big) = (1 + 4\varepsilon_n^2)^n.$$

Since $4n\varepsilon_n^2 = \eta^2/64$,

$$(1 + 4\varepsilon_n^2)^n \le \exp(\eta^2/64) \le 1 + \eta^2/64 + (\eta^2/64)^2 \le 1 + \chi_\alpha.$$

The two-point length inequality above, with $\chi = \chi_\alpha$, yields

$$\max\left\{ \left(\eta - \frac{1}{2}\sqrt{\chi_\alpha} \right) \varepsilon_n, 0 \right\} \le L_n^\star.$$

Thus, for some $c_{\mathrm{root}} > 0$,

$$\max\left\{ c_{\mathrm{root}} n^{-1/2}, 0 \right\} \le L_n^\star$$

eventually.

For the local component, use the fixed compactly supported C^∞ bump $\mathfrak{b} : \mathbb{R} \to \mathbb{R}$ from the same two-point construction, normalized by

$$\mathfrak{b}(0) = 1, \qquad 0 \le \mathfrak{b}(u) \le 1, \qquad \mathfrak{b}(u) = 0 \text{ whenever } |u| \ge 1.$$

The smoothness calibration for this particular bump gives an amplitude $\mathfrak{a}_{\mathrm{bump}} \in (0, 1/4]$, depending only on (β, L), with the following uniform property. For every $0 \le \delta \le 1$ and $0 < h \le 1$, define

$$q_{\delta,h}^{\mathrm{bump}}(a) = \mathfrak{a}_{\mathrm{bump}} h^\beta \mathfrak{b}\left(\frac{a - \delta}{h} \right).$$

Then $a \mapsto 1/2 + q_{\delta,h}^{\mathrm{bump}}(a)$ is continuous on $[0, 1]$, takes values in $[0, 1]$, and satisfies the Hölder Taylor remainder

$$\left| \left(\frac{1}{2} + q_{\delta,h}^{\mathrm{bump}}(t) \right) - \sum_{j=0}^{\ell} \frac{\partial_{[0,1]}^j (1/2 + q_{\delta,h}^{\mathrm{bump}})(s)}{j!} (t - s)^j \right| \le L|t - s|^\beta$$

for all $s, t \in [0, 1]$. Indeed, applying the normalized Hölder gate of $\mathfrak{b}$ with radius $\min\{L, 1\}$ and then shrinking the amplitude to be at most $1/4$ gives both the displayed Taylor bound and the range constraint.

Set

$$\zeta_{\mathrm{loc}} = \frac{\eta^2/64}{8(\kappa + 1)\mathfrak{a}_{\mathrm{bump}}^2 + 1}, \qquad h_{n,\mathrm{loc}} = \zeta_{\mathrm{loc}} h_n.$$

Let $\mathbb{P}_{0,n}^{\mathrm{loc}}$ and $\mathbb{P}_{1,n}^{\mathrm{loc}}$ have the same common design as the root pair, and let their conditional Bernoulli success probabilities be $1/2 + g_{j,n}^{\mathrm{loc}}(x, a)$, where

$$g_{0,n}^{\mathrm{loc}}(x, a) = 0, \qquad g_{1,n}^{\mathrm{loc}}(x, a) = \mathfrak{a}_{\mathrm{bump}} h_{n,\mathrm{loc}}^\beta \mathfrak{b}\left(\frac{a - \delta_n}{h_{n,\mathrm{loc}}} \right).$$

For all sufficiently large n, Lemma 10 gives $0 < h_n \le 1$, hence $0 < h_{n,\mathrm{loc}} \le 1$, and the amplitude calibration places both laws in $\mathcal{M}$. The localized shift is nonnegative and equals $\mathfrak{a}_{\mathrm{bump}} h_{n,\mathrm{loc}}^\beta$ at $a = \delta_n$, so the collapsed atom contribution alone gives

$$\left| \theta_{\delta_n}(\mathbb{P}_{1,n}^{\mathrm{loc}}) - \theta_{\delta_n}(\mathbb{P}_{0,n}^{\mathrm{loc}}) \right| \ge \mathfrak{a}_{\mathrm{bump}} \delta_n^{\kappa+1} h_{n,\mathrm{loc}}^\beta = \mathfrak{a}_{\mathrm{bump}} \zeta_{\mathrm{loc}}^\beta \delta_n^{\kappa+1} h_n^\beta.$$

For one observation, write

$$\mathcal{I}_n = \int 4\{g^{\mathrm{loc}}_{1,n}(x,a)\}^2 \, d\mathbb{P}_{X,A}(x,a),$$

where $\mathbb{P}_{X,A}$ is the common design law. Since $0 \le \mathfrak{b} \le 1$ and $\mathfrak{b}$ vanishes outside $[-1,1]$,

$$\mathcal{I}_n \le 8(\kappa+1)\mathfrak{a}^2_{\mathrm{bump}} h^{2\beta+1}_{n,\mathrm{loc}} (\delta_n + h_{n,\mathrm{loc}})^{\kappa}.$$

Because $0 < \zeta_{\mathrm{loc}} \le 1$, $h_{n,\mathrm{loc}} \le h_n$. With $2\beta+1 > 1$, $\kappa \ge 0$, and the balance identity from Lemma 10,

$$\begin{aligned} n\mathcal{I}_n &\le 8(\kappa+1)\mathfrak{a}^2_{\mathrm{bump}} \zeta^{2\beta+1}_{\mathrm{loc}} \, n h_n^{2\beta+1} (\delta_n + h_n)^{\kappa} \\ &\le \eta^2/64. \end{aligned}$$

Independence gives

$$1 + \chi^2\Big((\mathbb{P}^{\mathrm{loc}}_{1,n})^{\otimes n}, (\mathbb{P}^{\mathrm{loc}}_{0,n})^{\otimes n}\Big) = (1+\mathcal{I}_n)^n \le \exp(\eta^2/64) \le 1 + \chi_\alpha.$$

The two-point length inequality therefore gives, for some $c_{\mathrm{loc}} > 0$,

$$\max\Big\{c_{\mathrm{loc}} \delta_n^{\kappa+1} h_n^{\beta}, 0\Big\} \le L_n^{\star}$$

eventually. Taking half the smaller of c_{root} and c_{loc}, and splitting into the two cases according to which component is larger, gives a constant $c_{\min} > 0$ such that

$$\max\Big\{c_{\min}\left(n^{-1/2} + \delta_n^{\kappa+1} h_n^{\beta}\right), 0\Big\} \le L_n^{\star}$$

eventually.

The every-procedure version uses that a uniformly honest procedure is an admissible competitor in the minimax honest-length criterion. Hence there is a positive constant $c_{\mathrm{proc}} > 0$ such that every uniformly honest observed-sample interval procedure C_n satisfies

$$\max\Big\{c_{\mathrm{proc}}\left(n^{-1/2} + \delta_n^{\kappa+1} h_n^{\beta}\right), 0\Big\} \le \sup\left\{\int^{-} \mathrm{len}(C_n)\, dP^{\otimes n} : P \in \mathcal{M}\right\}$$

in $\mathbb{R}_{\ge 0} \cup \{\infty\}$.

3. Define
$$c = \frac{1}{2}\min\{c_{\min}, c_{\mathrm{proc}}\}, \qquad C = C_{\mathrm{up}} + c + 1.$$
Then $c > 0$, $c < C$, $c \le c_{\min}$, $c \le c_{\mathrm{proc}}$, and $C_{\mathrm{up}} \le C$.

4. Fix an admissible threshold path (δ_n) and split path $B_\bullet$. By Lemma 10, for all sufficiently large n,
$$0 < h_n, \qquad h_n \le 1 - \bar{\delta}, \qquad n h_n^{2\beta+1}(\delta_n + h_n)^{\kappa} = 1.$$
Together with $0 \le \delta_n \le \bar{\delta}$, this gives $\delta_n + h_n \le 1$. Hence the hypotheses of Lemma 14 hold. That result supplies observed-sample measurability of CI_n and the modelwise coverage inequality
$$1 - \alpha \le P^{\otimes n}\{\theta_{\delta_n}(P) \in \mathrm{CI}_n\}$$
for every $P \in \mathcal{M}$, so the stabilized interval is uniformly honest. Taking the infimum over P gives the displayed worst-case coverage bound.

5. Put
$$r_n = n^{-1/2} + \delta_n^{\kappa+1} h_n^{\beta}.$$
Since $n^{-1/2} > 0$ and $\delta_n^{\kappa+1} h_n^{\beta} \geq 0$, we have $r_n > 0$. The upper ingredient and $C_{\mathrm{up}} \leq C$ yield
$$\sup_{P \in \mathcal{M}} \mathbb{E}_{P^{\otimes n}}[\mathrm{len}(\mathrm{CI}_n)] \leq C r_n.$$
The minimax lower ingredient and $c \leq c_{\min}$ yield the extended-real comparison
$$\max\{c r_n, 0\} \leq L_n^{\star}.$$

Since CI_n is itself uniformly honest by the coverage step, the procedure lower bound applies to this concrete interval. Combined with $c \leq c_{\mathrm{proc}}$, it gives
$$\max\{c r_n, 0\} \leq \sup \left\{ \int^{-} \mathrm{len}(\mathrm{CI}_n) \, dP^{\otimes n} : P \in \mathcal{M} \right\}$$
in $[0, \infty]$. For the concrete stabilized interval the length is measurable, nonnegative, and bounded by one, so its worst-case length in $[0, \infty]$ agrees with its ordinary worst-case expected length, read in $[0, \infty]$ as in Definition 23. Since $c r_n > 0$, this gives the real lower bound
$$c r_n \leq \sup_{P \in \mathcal{M}} \mathbb{E}_{P^{\otimes n}}[\mathrm{len}(\mathrm{CI}_n)] \, .$$
Finally, for every uniformly honest observed-sample interval procedure C_n, the procedure lower bound gives the extended-real statement
$$\max\{c r_n, 0\} \leq \sup \left\{ \int^{-} \mathrm{len}(C_n) \, dP^{\otimes n} : P \in \mathcal{M} \right\},$$
with the supremum taken in $\mathbb{R}_{\geq 0} \cup \{\infty\}$. This is the arbitrary-procedure worst-length criterion used in the comparison. For interval procedures whose pointwise lengths lie in $[0, 1]$ as in Definition 15, the nonnegative lower integral equals the ordinary expected length under the positive-part embedding of Definition 23, and the display specializes to the corresponding real expected-length lower bound. These are the eventual length comparisons established by the argument.

□

Proof of Theorem 3. Let
$$\Gamma = \beta\kappa + 2\beta + \kappa + 1, \qquad \Xi = 2\beta + \kappa + 1, \qquad \zeta = \frac{\Gamma}{2\beta + 1}.$$
The regime assumptions give $\Gamma > 0$, $\Xi > 0$, and $\zeta > 0$.

1. The two deterministic scales satisfy
$$\frac{\delta_{\mathrm{crit},n}}{\delta_{\mathrm{edge},n}} = \frac{n^{-1/(2\Gamma)}}{n^{-1/\Xi}} = n^{1/\Xi - 1/(2\Gamma)}$$
for every $n \geq 1$. The exponent is positive because
$$2\Gamma - \Xi = 2\beta\kappa + 2\beta + \kappa + 1 > 0.$$

Hence $n^{1/\Xi-1/(2\Gamma)} \to \infty$, which proves

$$\frac{\delta_{\mathrm{crit},n}}{\delta_{\mathrm{edge},n}} \to \infty.$$

2. Fix a deterministic threshold sequence (δ_n) with $\delta_n \in [0, \bar{\delta}]$. Let h_n be the information-balance bandwidth from Definition 8, define

$$a_n = \delta_n^{\kappa+1} h_n^{\beta},$$

and let r_n be the frontier rate in Definition 9, so that $r_n = n^{-1/2} + a_n$. By Lemma 10, for all sufficiently large n,

$$h_n > 0, \qquad h_n \le 1 - \bar{\delta}, \qquad n h_n^{2\beta+1}(\delta_n + h_n)^{\kappa} = 1.$$

The phase comparisons below use this deterministic balance relation directly. In the regular regime $\delta_n/\delta_{\mathrm{crit},n} \to 0$, we also have, for all sufficiently large n,

$$\frac{\delta_n}{\delta_{\mathrm{crit},n}} < 1.$$

If $\delta_n = 0$, then $a_n = 0$ because $\kappa + 1 > 0$. If $\delta_n > 0$, set

$$\eta_n = \left(n\delta_n^{\kappa}\right)^{-1/(2\beta+1)}.$$

Since $\delta_n + h_n \ge \delta_n$ and $\kappa \ge 0$, the balance equation implies

$$h_n \le \eta_n.$$

Therefore

$$a_n = \delta_n^{\kappa+1} h_n^{\beta} \le \delta_n^{\kappa+1}\eta_n^{\beta}.$$

The exact power computation is

$$\frac{\delta_n^{\kappa+1}\eta_n^{\beta}}{n^{-1/2}} = \left(\frac{\delta_n}{\delta_{\mathrm{crit},n}}\right)^{\zeta} \le 1.$$

Thus $0 \le a_n \le n^{-1/2}$ eventually, and

$$n^{-1/2} \le r_n = n^{-1/2} + a_n \le 2n^{-1/2}$$

eventually. Hence $r_n \asymp n^{-1/2}$.

3. Suppose $\delta_n/\delta_{\mathrm{crit},n} \to c_0$ with $c_0 > 0$. By the scale separation $\delta_{\mathrm{crit},n}/\delta_{\mathrm{edge},n} \to \infty$,

$$\frac{\delta_n}{\delta_{\mathrm{edge},n}} = \frac{\delta_n}{\delta_{\mathrm{crit},n}}\frac{\delta_{\mathrm{crit},n}}{\delta_{\mathrm{edge},n}} \to \infty.$$

For this above-edge regime, the balance equation gives the closed-form bandwidth order. Indeed, for all sufficiently large n, $\delta_n > 0$ and $\delta_{\mathrm{edge},n} \le \delta_n$. With $\eta_n = (n\delta_n^{\kappa})^{-1/(2\beta+1)}$, the balance equation gives $h_n \le \eta_n$. The inequality $\delta_{\text{edge},n} \le \delta_n$ gives $\eta_n \le \delta_n$, and hence $\delta_n + h_n \le 2\delta_n$. Substituting this upper bound into the same balance equation yields

$$2^{-\kappa/(2\beta+1)}\eta_n \le h_n \le \eta_n.$$

Consequently

$$h_n \asymp \left(n\delta_n^\kappa\right)^{-1/(2\beta+1)}.$$

Multiplying by the nonnegative factor $\delta_n^{\kappa+1}$ and raising the bandwidth comparison to the positive power β gives

$$a_n \asymp \delta_n^{\kappa+1}\left\{\left(n\delta_n^\kappa\right)^{-1/(2\beta+1)}\right\}^\beta.$$

The normalized interior term satisfies

$$\frac{\delta_n^{\kappa+1}\left\{\left(n\delta_n^\kappa\right)^{-1/(2\beta+1)}\right\}^\beta}{n^{-1/2}} = \left(\frac{\delta_n}{\delta_{\text{crit},n}}\right)^\zeta \to c_0^\zeta \in (0,\infty).$$

Consequently $a_n \asymp n^{-1/2}$. Since $r_n = n^{-1/2} + a_n$ and both summands are nonnegative, also

$$r_n \asymp n^{-1/2}.$$

4. Suppose $\delta_n/\delta_{\text{crit},n} \to \infty$ and $\delta_n \to 0$. The scale separation $\delta_{\text{crit},n}/\delta_{\text{edge},n} \to \infty$ again yields

$$\frac{\delta_n}{\delta_{\text{edge},n}} \to \infty,$$

so the same direct balance comparison used at the critical scale gives

$$h_n \asymp \left(n\delta_n^\kappa\right)^{-1/(2\beta+1)}.$$

Thus

$$a_n \asymp \delta_n^{\kappa+1}\left(n\delta_n^\kappa\right)^{-\beta/(2\beta+1)}.$$

Moreover,

$$\frac{\delta_n^{\kappa+1}\left(n\delta_n^\kappa\right)^{-\beta/(2\beta+1)}}{n^{-1/2}} = \left(\frac{\delta_n}{\delta_{\text{crit},n}}\right)^\zeta \to \infty.$$

Hence the root term is eventually bounded by a constant multiple of the displayed atom scale, and

$$r_n = n^{-1/2} + a_n \asymp \delta_n^{\kappa+1}\left(n\delta_n^\kappa\right)^{-\beta/(2\beta+1)}.$$

5. Suppose $\delta_n \to \delta_0$ with $\delta_0 \in (0, \bar{\delta}]$. Since

$$\delta_{\text{crit},n} = n^{-1/(2\Gamma)} \to 0,$$

we have

$$\frac{\delta_n}{\delta_{\text{crit},n}} \to \infty.$$

Together with the scale separation $\delta_{\text{crit},n}/\delta_{\text{edge},n} \to \infty$, this places the sequence above the edge scale. The same direct balance comparison used in the supercritical case gives

$$r_n \asymp \delta_n^{\kappa+1}\left(n\delta_n^\kappa\right)^{-\beta/(2\beta+1)}.$$

For all sufficiently large n, $\delta_n > 0$, and

$$\delta_n^{\kappa+1}\big(n\delta_n^{\kappa}\big)^{-\beta/(2\beta+1)} = n^{-\beta/(2\beta+1)}\delta_n^{\zeta}.$$

Because $\delta_n^{\zeta} \to \delta_0^{\zeta} \in (0,\infty)$, it follows that

$$r_n \asymp n^{-\beta/(2\beta+1)}.$$

6. Fix n and $P \in \mathcal{M}$. At threshold 0, Definition 7 gives, for every stratum x,

$$q_x(0) = \int_{[0,0]} \pi_x(a)\,\mathrm{d}a = 0.$$

The conditional-density condition in Assumption 2 gives

$$P(A=0, X=x) = p_x \int_{\{0\}} \pi_x(a)\,\mathrm{d}a = 0$$

for every stratum x. Summing over the finite stratum set yields $P(A=0)=0$, and the treatment support encoded by Definition 5 gives $A>0$ almost surely. Therefore the retained term in Definition 7 satisfies

$$\nu_0(P) = \int Y\,\mathbf{1}\{A>0\}\,dP = \int Y\,dP.$$

Since every $q_x(0)$ is zero,

$$\theta_0(P) = \nu_0(P) + \sum_x p_x q_x(0)\mu_x^P(0) = \int Y\,dP.$$

Finally, $\kappa+1>0$, so Definition 9 gives

$$r_n\big|_{\delta_n=0} = n^{-1/2} + 0^{\kappa+1}h_n^{\beta} = n^{-1/2}.$$

7. Fix n, $P \in \mathcal{M}$, and an admissible split $B=(I_0,I_1,I_2)$ as in Definition 1. The conditional-density condition in Assumption 2 gives $P(A=0)=0$, so $A>0$ under P almost surely. Since the product law in Definition 2 has each observed coordinate distributed as P, every coordinate satisfies $A_i>0$ under $P^{\otimes n}$ almost surely, and the finite intersection of these coordinate events still has probability one. On that event, the finite-block average is interpreted as the totalized expression

$$\mathrm{Avg}_I(Y) = \begin{cases} |I|^{-1}\sum_{i\in I} Y_i, & |I|>0,\\ 0, & |I|=0,\end{cases}$$

which is the displayed product $(|I|)^{-1}\sum_{i\in I} Y_i$ with the real-inverse convention $0^{-1}=0$. Thus

$$\widehat{b}_x = |I_1|^{-1}\sum_{i\in I_1}\mathbf{1}\{X_i=x,\ A_i\le 0\} = 0 \qquad \text{for every } x,$$

while

$$\widehat{\nu}_n = |I_0|^{-1}\sum_{i\in I_0} Y_i.$$

Substituting these identities into Algorithm 1 gives, $P^{\otimes n}$-almost surely,

$$\widehat{\theta}_n = \Pi_{[0,1]}\left(|I_0|^{-1}\sum_{i\in I_0} Y_i + \sum_x 0\cdot \widetilde{\mu}_x\right) = \Pi_{[0,1]}\left(|I_0|^{-1}\sum_{i\in I_0} Y_i\right).$$

Thus the threshold-zero total-Gram estimator agrees under the product sampling law with the clamped block-zero outcome average.

8. Apply Theorem 2 to the deterministic sequence $\delta_n \equiv 0$ and to a deterministic admissible split sequence $B_\bullet = (B_n)_{n\geq 1}$. The identity $r_n = n^{-1/2}$ at threshold zero gives eventual worst-case coverage
$$\inf_{P\in\mathcal{M}} P^{\otimes n}\{\theta_0(P) \in \mathrm{CI}_{n,B_n}\} \geq 1-\alpha,$$
and constants $0 < c_{\mathrm{len}} \leq C_{\mathrm{len}} < \infty$ such that, eventually, the concrete stabilized length from Definition 16 satisfies
$$c_{\mathrm{len}}\, n^{-1/2} \leq \mathrm{Length}_n(B_n) \leq C_{\mathrm{len}}\, n^{-1/2}.$$
Likewise, applying Theorem 1 to the same zero sequence and using $r_n = n^{-1/2}$ gives constants $0 < c_{\mathrm{risk}} \leq C_{\mathrm{risk}} < \infty$ such that, eventually, the concrete stabilized risk from Definition 16 satisfies
$$c_{\mathrm{risk}}\, n^{-1/2} \leq \mathrm{Risk}_n(B_n) \leq C_{\mathrm{risk}}\, n^{-1/2}.$$
Therefore the stabilized zero-threshold procedure has coverage at least $1-\alpha$ for all sufficiently large n, and its stabilized worst-case absolute-error risk and stabilized worst-case expected length at threshold 0 are both $\asymp n^{-1/2}$.

□

Proof of Proposition 1. 1. Put

$$\tau_{\delta,h}(a) = \max\{0, 1-|a-\delta|/h\}, \qquad \gamma_{\delta,h}(a) = h\,\tau_{\delta,h}(a),$$

so that, on the eventual range where $h > 0$,

$$m_0(a) = \frac{1}{2}, \qquad m_{1,\delta,h}(a) = \frac{1}{2} + \gamma_{\delta,h}(a), \qquad \pi(a) = 2a.$$

Let
$$\zeta_n = ((n)\delta_n)^{-1/3}.$$
For the infimum construction in Definition 8, set

$$F_n(h) = nh^3(\delta_n + h).$$

The bounds $0 \leq \delta_n \leq \bar{\delta}$ and $0 < \bar{\delta} < 1$ make F_n continuous and strictly increasing on $[0, 1-\bar{\delta}]$, with $F_n(0) = 0$. For all sufficiently large n, $F_n(1-\bar{\delta}) \geq 1$, so the crossing set in Definition 8 is a nonempty interval whose left endpoint is the positive solution of $F_n(h) = 1$. Hence, for all sufficiently large n,

$$0 < h_n \leq 1-\bar{\delta}, \qquad nh_n^3(\delta_n + h_n) = 1.$$

Since $\delta_n/n^{-1/4} \to \infty$, eventually $n^{-1/4} \le \delta_n$, hence $0 < \delta_n$ and

$$n\delta_n^4 \ge 1.$$

The balance equation gives

$$1 = nh_n^3(\delta_n + h_n) \ge nh_n^3\delta_n,$$

and therefore $h_n \le \zeta_n$. The inequality $n\delta_n^4 \ge 1$ is exactly

$$\zeta_n = (n\delta_n)^{-1/3} \le \delta_n.$$

Consequently $h_n \le \delta_n$, and then

$$1 = nh_n^3(\delta_n + h_n) \le 2n\delta_n h_n^3,$$

so

$$2^{-1/3}\zeta_n \le h_n \le \zeta_n$$

eventually. Also $\delta_n + h_n \le \bar{\delta} + (1 - \bar{\delta}) = 1$, and the same balance equation forces $h_n \le 1/4$ eventually, since $h_n > 1/4$ and $n \ge 257$ would imply

$$1 = nh_n^3(\delta_n + h_n) \ge nh_n^4 > 1.$$

Thus

$$h_n \asymp (n\delta_n)^{-1/3}.$$

This is the one-stratum $\beta = \kappa = 1$ far-edge instance of the bandwidth comparison in Lemma 16, specialized to the density $\pi(a) = 2a$ and the condition $\delta_n/n^{-1/4} \to \infty$.

2. For all sufficiently large n, the support of γ_{δ_n,h_n} is contained in $[\delta_n - h_n, \delta_n + h_n] \subset [0, 1]$. On the left and right halves of this interval,

$$\gamma_{\delta_n,h_n}(a) = a - \delta_n + h_n \quad (\delta_n - h_n \le a \le \delta_n),$$

and

$$\gamma_{\delta_n,h_n}(a) = \delta_n + h_n - a \quad (\delta_n \le a \le \delta_n + h_n).$$

Hence

$$\begin{aligned}\int_0^1 \gamma_{\delta_n,h_n}(a)^2\, 2a\, da &= \int_0^{h_n} 2(\delta_n - u)(h_n - u)^2\, du + \int_0^{h_n} 2(\delta_n + u)(h_n - u)^2\, du \\ &= 4\delta_n \int_0^{h_n} (h_n - u)^2\, du = \frac{4}{3}\,\delta_n h_n^3.\end{aligned}$$

3. Because $0 \le \tau_{\delta_n,h_n} \le 1$ and $h_n \le 1/4$ eventually,

$$|\gamma_{\delta_n,h_n}(a)| \le \frac{1}{4} \qquad (0 \le a \le 1).$$

For every real g with $|g| \le 1/4$, the Bernoulli comparison used here is

$$2g^2 \le \mathrm{KL}\left(\mathrm{Bernoulli}\left(\frac{1}{2} + g\right), \mathrm{Bernoulli}\left(\frac{1}{2}\right)\right) \le 4g^2.$$

The lower bound follows from Pinsker applied to the one-point event $\{1\}$, and the upper bound follows from the exact Bernoulli KL formula on the interval $[1/4, 3/4]$. Integrating this display against the nonnegative density $2a$ and using the local perturbation-energy computation gives

$$\frac{8}{3}\, n\delta_n h_n^3 \le K_n \le \frac{16}{3}\, n\delta_n h_n^3$$

eventually. Therefore

$$K_n \asymp n\delta_n h_n^3.$$

4. The collapsed atom mass is

$$\int_0^{\delta_n} 2a\, da = \delta_n^2,$$

and $\gamma_{\delta_n,h_n}(\delta_n) = h_n$. The retained-course contribution is supported on $(\delta_n, \delta_n + h_n]$, so

$$\begin{aligned}
\int_{\delta_n}^1 \gamma_{\delta_n,h_n}(a)\, 2a\, da &= \int_{\delta_n}^{\delta_n+h_n} (\delta_n + h_n - a)\, 2a\, da \\
&= \int_0^{h_n} 2(\delta_n + u)(h_n - u)\, du = \delta_n h_n^2 + \frac{1}{3} h_n^3.
\end{aligned}$$

Thus

$$\Delta_n = \delta_n^2 h_n + \delta_n h_n^2 + \frac{1}{3} h_n^3.$$

Since $0 < h_n \le \delta_n$ eventually,

$$\delta_n^2 h_n \le \Delta_n \le \frac{7}{3}\, \delta_n^2 h_n$$

eventually, and hence

$$\Delta_n \asymp \delta_n^2 h_n.$$

5. Let

$$\chi_n = n^{-1/10}, \qquad \varrho_n = n^{-1/2}, \qquad \upsilon_n = \delta_n^2 \zeta_n = \delta_n^2 (n\delta_n)^{-1/3}.$$

For all sufficiently large n, $\delta_n > 0$, and direct exponent algebra gives

$$\frac{\upsilon_n}{\varrho_n} = \left(\frac{\delta_n}{\chi_n}\right)^{5/3}.$$

Combining the bandwidth comparison from the first item with the target comparison from the fourth item gives constants $\underline{\mathfrak{h}}, \overline{\mathfrak{h}} > 0$ such that eventually

$$\underline{\mathfrak{h}}\, \upsilon_n \le \Delta_n \le \overline{\mathfrak{h}}\, \upsilon_n.$$

If $\delta_n \asymp \chi_n$, then for some $\underline{\mathfrak{d}}, \overline{\mathfrak{d}} > 0$,

$$\underline{\mathfrak{d}} \le \frac{\delta_n}{\chi_n} \le \overline{\mathfrak{d}}$$

eventually. Since $x \mapsto x^{5/3}$ is increasing on $[0, \infty)$,

$$\underline{\mathfrak{h}}\, \underline{\mathfrak{d}}^{5/3} \varrho_n \le \Delta_n \le \overline{\mathfrak{h}}\, \overline{\mathfrak{d}}^{5/3} \varrho_n$$

eventually, so $\Delta_n \asymp n^{-1/2}$.

Conversely, assume $\Delta_n \asymp \varrho_n$. Then for some $\underline{\mathfrak{s}}, \bar{\mathfrak{s}} > 0$,

$$\underline{\mathfrak{s}}\varrho_n \le \Delta_n \le \bar{\mathfrak{s}}\varrho_n$$

eventually. With

$$\lambda_{\mathrm{lo}} = \left(\frac{\underline{\mathfrak{s}}}{\bar{\mathfrak{h}}}\right)^{3/5}, \qquad \lambda_{\mathrm{hi}} = \left(\frac{\bar{\mathfrak{s}}}{\underline{\mathfrak{h}}}\right)^{3/5},$$

the two-sided display for Δ_n, together with $\upsilon_n/\varrho_n = (\delta_n/\chi_n)^{5/3}$, yields

$$\lambda_{\mathrm{lo}} \le \frac{\delta_n}{\chi_n} \le \lambda_{\mathrm{hi}}$$

eventually. Hence

$$\lambda_{\mathrm{lo}}\chi_n \le \delta_n \le (\lambda_{\mathrm{lo}} + \lambda_{\mathrm{hi}})\chi_n$$

eventually, which is $\delta_n \asymp n^{-1/10}$. This proves

$$\delta_n \asymp n^{-1/10} \quad \iff \quad \Delta_n \asymp n^{-1/2}.$$

□

Proof of Proposition 2. Stratum product identity and the cell formula. Fix $x \in \mathcal{X}$. From the regime constants and Definition 20,

$$p_x = P(X = x) = P^{\mathrm{F}}(X = x) \ge p_{\min} > 0, \qquad 0 \le \kappa, \qquad 0 \le c_+, \qquad \bar{\delta} \le 1.$$

Let η_x be the latent-coordinate marginal of the normalized full-data stratum law $p_x^{-1}P^{\mathrm{F}}|_{\{X=x\}}$. The exchangeability member of Definition 20, together with Assumption 2 for the observed margin, gives the product formula

$$p_x^{-1}P^{\mathrm{F}}\{X = x,\ A \in da,\ U \in du\} = \pi_x(a)\mathbf{1}_{[0,1]}(a)\, da\, \eta_x(du).$$

Consequently, for every Borel $B \subseteq [0, 1]$, consistency on the common full-data event gives

$$\begin{aligned} p_x^{-1}\int_{\{X=x,\ A\in B\}} Y\, dP^{\mathrm{F}} &= \int_B \int_{\mathcal{U}} g_x(a, u)\, \eta_x(du)\, \pi_x(a)\, da \\ &= \int_B m_x^{\mathrm{F}}(a)\pi_x(a)\, da. \end{aligned}$$

The measurability, boundedness, and restriction-to-$[0, 1]$ bookkeeping in this calculation is supplied by the jointly measurable structural response, the simultaneous consistency event, and the bounded $[0, 1]$-valued response range in Definition 20.

Pointwise identification of the threshold-range regression. For every Borel $B_\circ \subseteq [0, \bar{\delta}]$, the stratum product formula for $p_x^{-1}P^{\mathrm{F}}$ established for this stratum and equality of the observed margin give

$$\int_{B_\circ} m_x^{\mathrm{F}}(a)\pi_x(a)\, da = p_x^{-1}\int_{\{X=x,\ A\in B_\circ\}} Y\, dP = \int_{B_\circ} \mu_x^P(a)\pi_x(a)\, da,$$

where the second equality is the conditional-expectation version property in Assumption 5 combined with Assumption 2. The products are integrable on $[0, \bar{\delta}]$: both regression curves are continuous there, and the polynomial envelope in Assumption 4 bounds $\pi_x(a)$ by $c_+ a^\kappa$ on $[0,1]$. Hence

$$\int_{B_\circ} \{m_x^{\mathrm{F}}(a) - \mu_x^P(a)\}\pi_x(a)\,da = 0 \qquad \text{for all Borel } B_\circ \subseteq [0,\bar{\delta}].$$

Since $c_- > 0$ and $\kappa \geq 0$, the same thinning envelope gives

$$\pi_x(a) > 0 \qquad \text{for Lebesgue-a.e. } a \in (0,\bar{\delta}].$$

Thus $m_x^{\mathrm{F}} = \mu_x^P$ Lebesgue-a.e. on $(0,\bar{\delta}]$. Both functions are continuous on $[0,\bar{\delta}]$, so equality extends to every $a \in [0,\bar{\delta}]$:

$$m_x^{\mathrm{F}}(a) = \mu_x^P(a).$$

This also covers the endpoint $a = 0$ by right-continuity on the closed threshold interval.

Support preservation. Fix $\delta \in [0,\bar{\delta}]$. The observed margin P is in $\mathcal{M}$, and the regime constants satisfy the hypotheses of Proposition 7. Therefore, for each $x \in \mathcal{X}$,

$$d_\delta(a) \in \operatorname{supp}\{\pi_x(a)\,da \text{ on } [0,1]\} \qquad \text{for conditional-treatment-a.e. } a.$$

Pathwise clamp decomposition. Because $\delta \in [0,\bar{\delta}] \subseteq [0,1]$, the common consistency event applies both at A and at δ. On that event, if $\delta < A$, then $d_\delta(A) = A$ and $Y(d_\delta(A)) = Y$; if $A \leq \delta$, then $d_\delta(A) = \delta$ and $Y(d_\delta(A)) = Y(\delta)$. Hence

$$Y(d_\delta(A)) = Y\mathbf{1}\{\delta < A\} + Y(\delta)\mathbf{1}\{A \leq \delta\} \qquad P^{\mathrm{F}}\text{-a.s.}$$

The atom contribution. Define, for a full-data state ω,

$$\varphi_\delta^{>}(\omega) = Y(\omega)\mathbf{1}\{\delta < A(\omega)\}, \qquad \varphi_\delta^{\leq}(\omega) = Y(\delta)(\omega)\mathbf{1}\{A(\omega) \leq \delta\}.$$

The bounded outcome and potential-outcome range make these integrable. The observed-margin identity gives

$$\int \varphi_\delta^{>}\,dP^{\mathrm{F}} = \int Y\mathbf{1}\{\delta < A\}\,dP = \nu_\delta(P).$$

For the lower-tail term, partition over the finite strata and use the stratum product formula for $p_x^{-1}P^{\mathrm{F}}$ at the fixed dose δ:

$$\begin{aligned}
\int \varphi_\delta^{\leq}\,dP^{\mathrm{F}} &= \sum_{x\in\mathcal{X}} \int_{\{X=x,\ A\leq\delta\}} Y(\delta)\,dP^{\mathrm{F}} \\
&= \sum_{x\in\mathcal{X}} p_x \left(\int_0^\delta \pi_x(a)\,da\right) m_x^{\mathrm{F}}(\delta) \\
&= \sum_{x\in\mathcal{X}} p_x q_x(\delta) m_x^{\mathrm{F}}(\delta).
\end{aligned}$$

Target equality. Integrating the pathwise clamp decomposition of $Y(d_\delta(A))$ and substituting the two integrated clamp contributions $\int \varphi_\delta^{>}\,dP^{\mathrm{F}}$ and $\int \varphi_\delta^{\leq}\,dP^{\mathrm{F}}$ yields

$$\psi_\delta(P^{\mathrm{F}}) = \nu_\delta(P) + \sum_{x\in\mathcal{X}} p_x q_x(\delta) m_x^{\mathrm{F}}(\delta).$$

Since $\delta \in [0, \bar{\delta}]$, the identification $m_x^{\mathrm{F}} = \mu_x^P$ on $[0, \bar{\delta}]$ gives $m_x^{\mathrm{F}}(\delta) = \mu_x^P(\delta)$ for every stratum, and therefore

$$\psi_\delta(P^{\mathrm{F}}) = \nu_\delta(P) + \sum_{x \in \mathcal{X}} p_x q_x(\delta) \mu_x^P(\delta) = \theta_\delta(P).$$

All endpoint cases in the closed range $[0, \bar{\delta}]$, including $\delta = 0$ and $\delta = \bar{\delta}$, are covered by the same closed-interval hypotheses used above. $\square$

Proof of Proposition 3. 1. The regime assumptions give $J > 0$, $p_{\min} > 0$, and $\bar{\delta} \le 1$. Fix $P \in \mathcal{M}$. Define the total regression extension

$$m(x, a) = \begin{cases} \mu_x^P(a), & a \in [0, 1], \\ 0, & a \notin [0, 1]. \end{cases}$$

The model assumptions in Definition 5 give that m is Borel measurable, $m(x, a) \in [0, 1]$ for every (x, a), and

$$\mathbb{E}_P[Y \mid X, A] = m(X, A) \qquad P\text{-a.s.}$$

They also give continuity of $a \mapsto m(x, a)$ on $[0, \bar{\delta}]$ for every x, since $\bar{\delta} \le 1$, and positivity of every stratum mass:

$$0 < p_{\min} \le p_x = P\{X = x\}.$$

Let $\zeta = (x, a) \in \mathcal{X} \times \mathbb{R}$, let

$$\mathsf{Z}(o) = (o_X, o_A), \qquad \nu = P \circ \mathsf{Z}^{-1},$$

By Kallenberg [2002, Theorem 6.3], applied to the standard Borel design and outcome spaces, let K_ζ^{raw} be a conditional-distribution kernel of Y given $\mathsf{Z} = \zeta$, so that for all Borel $\mathsf{B} \subseteq \mathcal{X} \times \mathbb{R}$ and $\mathsf{C} \subseteq \mathbb{R}$,

$$P\{\mathsf{Z} \in \mathsf{B},\ Y \in \mathsf{C}\} = \int_{\mathsf{B}} K_\zeta^{\mathrm{raw}}(\mathsf{C})\, d\nu(\zeta).$$

Put

$$\tau(y) = \min\{1, \max\{0, y\}\}, \qquad K_\zeta^0 = K_\zeta^{\mathrm{raw}} \circ \tau^{-1}, \qquad \eta(\zeta) = \int y\, K_\zeta^0(dy).$$

Since $Y \in [0, 1]$ P-a.s.,

$$K_\zeta^{\mathrm{raw}} = K_\zeta^0 \quad \text{for } \nu\text{-a.e. } \zeta,$$

and the conditional-expectation identity gives

$$\eta(\mathsf{Z}) = m(\mathsf{Z}) \qquad P\text{-a.s.},$$

equivalently $\eta = m$ ν-a.e. Hence the Borel set

$$E = \{\zeta : \eta(\zeta) \ne m(\zeta)\}$$

satisfies $\nu(E) = 0$. Define the repaired kernel

$$K_\zeta^\star = \begin{cases} \delta_{m(\zeta)}, & \zeta \in E, \\ K_\zeta^0, & \zeta \in E^c. \end{cases}$$

Then $K^\star_\zeta$ is a probability law supported on $[0,1]$ for every ζ, $K^\star = K^{\mathrm{raw}}$ ν-a.e., and

$$\int y\, K^\star_\zeta(dy) = m(\zeta) \qquad \text{for every } \zeta.$$

Let $\lambda_{[0,1]}$ be Lebesgue probability measure on $[0,1]$. Realize the repaired kernel by its generalized inverse: for every ζ and $\upsilon \in [0,1]$, let $\mathfrak{q}(\zeta,\upsilon) \in [0,1]$ be

$$\mathfrak{q}(\zeta,\upsilon) = \sup\Big(\{0\} \cup \{s \in [0,1] : K^\star_\zeta([0,s]) < \upsilon\}\Big).$$

The strict superlevel sets of this supremum are countable unions over rational $s \in [0,1]$, so $(\zeta,\upsilon) \mapsto \mathfrak{q}(\zeta,\upsilon)$ is Borel. The right-continuity of $t \mapsto K^\star_\zeta((-\infty,t])$ gives the generalized-inverse identity, for every $t \in \mathbb{R}$,

$$\lambda_{[0,1]}\{\upsilon : \mathfrak{q}(\zeta,\upsilon) \le t\} = K^\star_\zeta((-\infty,t])$$

Therefore

$$\mathfrak{q}(\zeta,\cdot)_{\#}\lambda_{[0,1]} = K^\star_\zeta.$$

On the product space

$$(\mathcal{X} \times \mathbb{R}) \times [0,1], \qquad \nu \otimes \lambda_{[0,1]},$$

define

$$g_x(a,\upsilon) = \mathfrak{q}((x,a),\upsilon), \qquad Y(a) = g_X(a,U), \qquad Y = g_X(A,U).$$

Because $K^\star = K^{\mathrm{raw}}$ ν-a.e., the law of (X,A,Y) is exactly P. The displayed structural identities $Y(a) = g_X(a,U)$ and $Y = g_X(A,U)$ also give the simultaneous consistency condition in Definition 20. For bounded Borel $\varphi : \mathbb{R} \to \mathbb{R}$ and $\chi : [0,1] \to \mathbb{R}$, the product construction yields, for every x,

$$p_x \int_{\{X=x\}} \varphi(A)\chi(U)\, dP^{\mathrm{F}} = \left(\int_{\{X=x\}} \varphi(A)\, dP^{\mathrm{F}}\right)\left(\int_{\{X=x\}} \chi(U)\, dP^{\mathrm{F}}\right),$$

which is the exchangeability condition in Definition 20. Finally, for $a \in [0,\bar{\delta}]$,

$$m^{\mathrm{F}}_x(a) = p_x^{-1}\int_{\{X=x\}} g_x(a,U)\, dP^{\mathrm{F}} = \int_0^1 \mathfrak{q}((x,a),\upsilon)\, d\lambda_{[0,1]}(\upsilon) = \int y\, K^\star_{(x,a)}(dy) = m(x,a) = \mu^P_x(a).$$

Thus $a \mapsto m^{\mathrm{F}}_x(a)$ is continuous on $[0,\bar{\delta}]$. For the remainder of the proof, write

$$\Gamma(P^{\mathrm{F}}) := \text{the observed } (X,A,Y)\text{-margin of the full-data law package } P^{\mathrm{F}}.$$

The constructed package P^{F} belongs to $\mathcal{M}^{\mathrm{F}}$, and its observed margin is P. Hence

$$\forall P \in \mathcal{M} \quad \exists P^{\mathrm{F}} \in \mathcal{M}^{\mathrm{F}} \quad \text{with} \quad \Gamma(P^{\mathrm{F}}) = P.$$

This proves observed-margin surjectivity.

2. Fix $n \in \mathbb{N}$, including the empty-sample case $n = 0$, and fix $\delta \in [0, \bar{\delta}]$. For each $P^{\mathrm{F}} \in \mathcal{M}^{\mathrm{F}}$, Proposition 2 gives the target identity

$$\psi_\delta(P^{\mathrm{F}}) = \theta_\delta\{\Gamma(P^{\mathrm{F}})\}.$$

The causal criteria in Definition 22 evaluate every observed-sample estimator, coverage event, and length objective under the product law $\Gamma(P^{\mathrm{F}})^{\otimes n}$ of the observed margin, in the notation of Definition 2. Hence, after the displayed target identity, each causal risk, coverage probability, and expected length depends on P^{F} only through $\Gamma(P^{\mathrm{F}})$.

3. Let T_n be any observed-measurable estimator with values in $[0, 1]$. The target and sampling-law identity $\psi_\delta(P^{\mathrm{F}}) = \theta_\delta\{\Gamma(P^{\mathrm{F}})\}$ gives

$$\int \left|T_n(z) - \psi_\delta(P^{\mathrm{F}})\right| \,\mathrm{d}\Gamma(P^{\mathrm{F}})^{\otimes n}(z) = \int \left|T_n(z) - \theta_\delta\{\Gamma(P^{\mathrm{F}})\}\right| \,\mathrm{d}\Gamma(P^{\mathrm{F}})^{\otimes n}(z).$$

Taking suprema and using the surjectivity already proved,

$$\sup_{P^{\mathrm{F}} \in \mathcal{M}^{\mathrm{F}}} \int \left|T_n(z) - \psi_\delta(P^{\mathrm{F}})\right| \,\mathrm{d}\Gamma(P^{\mathrm{F}})^{\otimes n}(z) = \sup_{P \in \mathcal{M}} \int |T_n(z) - \theta_\delta(P)| \,\mathrm{d}P^{\otimes n}(z).$$

The admissible estimator class is the same on both sides, so taking the infimum over T_n and using Definitions 18 and 22 gives

$$R^\star_{n,\mathrm{F}}(\delta) = R^\star_n(\delta).$$

4. Let $\mathcal{C}_n$ be an observed-measurable interval procedure. The same target and product-law identities give

$$\Gamma(P^{\mathrm{F}})^{\otimes n}\{z : \psi_\delta(P^{\mathrm{F}}) \in \mathcal{C}_n(z)\} = \Gamma(P^{\mathrm{F}})^{\otimes n}\{z : \theta_\delta(\Gamma(P^{\mathrm{F}})) \in \mathcal{C}_n(z)\}.$$

Therefore, using surjectivity again,

$$\inf_{P^{\mathrm{F}} \in \mathcal{M}^{\mathrm{F}}} \Gamma(P^{\mathrm{F}})^{\otimes n}\{z : \psi_\delta(P^{\mathrm{F}}) \in \mathcal{C}_n(z)\} = \inf_{P \in \mathcal{M}} P^{\otimes n}\{z : \theta_\delta(P) \in \mathcal{C}_n(z)\}.$$

Thus the uniformly honest interval procedures are exactly the same for the causal and observed criteria. Their length objectives also agree:

$$\sup_{P^{\mathrm{F}} \in \mathcal{M}^{\mathrm{F}}} \int \operatorname{len}\{\mathcal{C}_n(z)\} \,\mathrm{d}\Gamma(P^{\mathrm{F}})^{\otimes n}(z) = \sup_{P \in \mathcal{M}} \int \operatorname{len}\{\mathcal{C}_n(z)\} \,\mathrm{d}P^{\otimes n}(z).$$

Taking the infimum over the common feasible class and using Definitions 18 and 22 gives

$$L^\star_{n,\mathrm{F}}(\delta, \alpha) = L^\star_n(\delta, \alpha).$$

□

Proof of Theorem 4. 1. For standard Borel spaces S, T and every probability law ϱ on $S \times T$, the usual disintegration gives a probability kernel $K_\varrho : S \rightsquigarrow T$ such that, for measurable $D \subseteq S$ and $E \subseteq T$,

$$\varrho(D \times E) = \int_D K_\varrho(s, E) \, d(\varrho \circ \mathrm{pr}_S^{-1})(s).$$

This standard Borel disintegration supplies the regular conditional law needed for the observed-margin comparison in Proposition 3. Let

$$\Gamma(P^{\mathrm{F}}) = \text{the observed } (X, A, Y)\text{-margin of } P^{\mathrm{F}}.$$

Under the regime conditions, Proposition 3 gives surjectivity of Γ from $\mathcal{M}^{\mathrm{F}}$ onto $\mathcal{M}$ and, for every n and every $\delta \in [0, \bar{\delta}]$,

$$R^{\star}_{n,\mathrm{F}} = R^{\star}_{n}, \qquad L^{\star}_{n,\mathrm{F}} = L^{\star}_{n}.$$

For every $P^{\mathrm{F}} \in \mathcal{M}^{\mathrm{F}}$ and every such δ, Proposition 2 gives the target identity

$$\psi_\delta(P^{\mathrm{F}}) = \theta_\delta\{\Gamma(P^{\mathrm{F}})\}.$$

2. Let $c_{\mathrm{risk}}, C_{\mathrm{risk}}$ be the constants supplied by Theorem 1, and let $c_{\mathrm{len}}, C_{\mathrm{len}}$ be the constants supplied by Theorem 2. Set

$$c = \min\{c_{\mathrm{risk}}, c_{\mathrm{len}}\}, \qquad C = \max\{C_{\mathrm{risk}}, C_{\mathrm{len}}\}.$$

Then $0 < c < C$. Choose $n_\star$ so that, for every $n \geq n_\star$, the eventual conclusions of Theorem 1, Theorem 2, and Lemma 10 hold, and also $n \geq 4$. Fix such an n, and write

$$\delta = \delta_n, \qquad h = h_n, \qquad r = r_n = n^{-1/2} + \delta^{\kappa+1} h^{\beta}.$$

Since $0 \leq \delta \leq \bar{\delta}$, Lemma 10 gives

$$0 < h, \qquad h \leq 1 - \bar{\delta}, \qquad n h^{2\beta+1} (\delta + h)^{\kappa} = 1,$$

and hence

$$\delta + h \leq \bar{\delta} + (1 - \bar{\delta}) = 1.$$

Moreover $r \geq 0$.

3. The total-Gram estimator in Algorithm 1 is a measurable function of the observed sample, so

$$z \longmapsto \widehat{\theta}_n(z)$$

is observed-sample measurable. The stabilized interval in Algorithm 2 has measurable endpoints built from the same finite sums and local-polynomial weights. The center is clipped to $[0, 1]$, and the displayed radius is nonnegative, so the clipped lower endpoint is no larger than the clipped upper endpoint. Thus CI_n is an observed-sample measurable interval.

4. For the concrete estimator, the definitions of the full-data and observed risks, the identity $\psi_\delta(P^{\mathrm{F}}) = \theta_\delta\{\Gamma(P^{\mathrm{F}})\}$, and the surjectivity of Γ give

$$\begin{aligned} &\sup_{P^{\mathrm{F}} \in \mathcal{M}^{\mathrm{F}}} \mathbb{E}_{\Gamma(P^{\mathrm{F}})^{\otimes n}} \left| \widehat{\theta}_n - \psi_\delta(P^{\mathrm{F}}) \right| \\ &\qquad = \sup_{P \in \mathcal{M}} \mathbb{E}_{P^{\otimes n}} \left| \widehat{\theta}_n - \theta_\delta(P) \right|. \end{aligned}$$

Combining this equality with Theorem 1 and the choices $c \leq c_{\mathrm{risk}}$, $C_{\mathrm{risk}} \leq C$ yields

$$c r \leq R^{\star}_{n,\mathrm{F}} \leq \sup_{P^{\mathrm{F}} \in \mathcal{M}^{\mathrm{F}}} \mathbb{E}_{\Gamma(P^{\mathrm{F}})^{\otimes n}} \left| \widehat{\theta}_n - \psi_\delta(P^{\mathrm{F}}) \right| \leq C r.$$

5. The same target transport gives the concrete coverage identity

$$\inf_{P^{\mathrm{F}}\in\mathcal{M}^{\mathrm{F}}} \Gamma(P^{\mathrm{F}})^{\otimes n}\{\psi_\delta(P^{\mathrm{F}}) \in \mathrm{CI}_n\} = \inf_{P\in\mathcal{M}} P^{\otimes n}\{\theta_\delta(P) \in \mathrm{CI}_n\}.$$

   The right-hand side is at least $1-\alpha$ by Theorem 2, so

$$\inf_{P^{\mathrm{F}}\in\mathcal{M}^{\mathrm{F}}} \Gamma(P^{\mathrm{F}})^{\otimes n}\{\psi_\delta(P^{\mathrm{F}}) \in \mathrm{CI}_n\} \geq 1-\alpha.$$

6. Expected length is transported by the same observed-margin argument. Since CI_n is a function of the observed sample, its length depends on P^{F} only through $\Gamma(P^{\mathrm{F}})$, and the surjectivity of Γ gives

$$\sup_{P^{\mathrm{F}}\in\mathcal{M}^{\mathrm{F}}} \mathbb{E}_{\Gamma(P^{\mathrm{F}})^{\otimes n}} \operatorname{len}(\mathrm{CI}_n) = \sup_{P\in\mathcal{M}} \mathbb{E}_{P^{\otimes n}} \operatorname{len}(\mathrm{CI}_n).$$

   For every observed law P, the clipping in Algorithm 2 gives

$$0 \leq \operatorname{len}(\mathrm{CI}_n(z)) \leq 1.$$

   Thus the real expected-length bound from Theorem 2 gives

$$\sup_{P\in\mathcal{M}} \mathbb{E}_{P^{\otimes n}} \operatorname{len}(\mathrm{CI}_n) \leq C_{\mathrm{len}} r \leq Cr,$$

   where the last inequality uses $C_{\mathrm{len}} \leq C$ and $r \geq 0$. Combining this bound with the equality of the causal and observed expected-length criteria gives

$$\sup_{P^{\mathrm{F}}\in\mathcal{M}^{\mathrm{F}}} \mathbb{E}_{\Gamma(P^{\mathrm{F}})^{\otimes n}} \operatorname{len}(\mathrm{CI}_n) \leq Cr.$$

   The same inequality may be viewed in $\mathbb{R}_{\geq 0} \cup \{\infty\}$ through the order-preserving embedding of finite nonnegative real numbers; this is the extended order used below for the minimax length criterion.

7. Finally, Lemma 14 makes CI_n an admissible honest observed-sample interval, since $n \geq 4$, $0 < h$, $\delta \in [0, \bar{\delta}]$, and $\delta + h \leq 1$. Hence, by the definition of $L_n^\star$ and the expected-length bound just proved,

$$L_{n,\mathrm{F}}^\star = L_n^\star \leq \sup_{P\in\mathcal{M}} \mathbb{E}_{P^{\otimes n}} \operatorname{len}(\mathrm{CI}_n) \leq Cr,$$

   with the two sides read in $\mathbb{R}_{\geq 0} \cup \{\infty\}$ by embedding the finite nonnegative quantities as finite extended values. The lower bound from Theorem 2, together with $c \leq c_{\mathrm{len}}$ and $r \geq 0$, gives, in the same extended order,

$$cr \leq c_{\mathrm{len}} r \leq L_n^\star = L_{n,\mathrm{F}}^\star.$$

   Substituting back $\delta = \delta_n$, $h = h_n$, and $r = r_n$ gives all displayed conclusions for every $n \geq n_\star$.

□

Proof of Proposition 4. Write P for the observed margin of P^{F}, and write $p_x = P(X = x) = P^{\mathrm{F}}(X = x)$. The hypotheses in Definition 27 and the design constants give

$$J > 0, \qquad \kappa \geq 0, \qquad c_- > 0, \qquad c_- \leq \kappa + 1 \leq c_+, \qquad 0 < p_{\min} \leq 1/J, \qquad 0 < \bar{\delta} < 1.$$

In particular, $p_x > 0$ for every stratum, $c_+ \geq 0$, and $\bar{\delta} \leq 1$.

1. Fix $x \in \mathcal{X}$. Let $\mu^P_{x,\mathrm{cont}}$ be the continuity-only regression extension selected in Definition 11; the observed-margin component of Definition 27 supplies the membership of P in $\mathcal{M}_{\mathrm{cont}}$. For every measurable $B \subseteq [0, \bar{\delta}]$, the full-data product identity from consistency, exchangeability, and the conditional-density law gives

$$\int_B m^{\mathrm{F}}_x(a)\pi_x(a)\, da = p_x^{-1} \int_{\{X=x,\ A\in B\}} Y\, dP^{\mathrm{F}}.$$

Indeed, under the normalized full-data law restricted to $X = x$, the treatment coordinate has the conditional law with density $\pi_x(a)\, da$ on $[0, 1]$. Its latent-coordinate marginal is the pushforward

$$\nu^U_x := \big((O, u) \mapsto u\big)_{\#} \left(p_x^{-1}\, P^{\mathrm{F}} \restriction \{(O, u) : X(O) = x\}\right),$$

equivalently, for every measurable $C \subseteq \mathcal{U}$,

$$\nu^U_x(C) = p_x^{-1} P^{\mathrm{F}}\{X = x,\ U \in C\}.$$

With this normalization, Assumptions 2, 6 and 7 give the product factorization of the treatment and latent coordinates in stratum x. Applying that factorization to $\mathbf{1}\{a \in B\}g_x(a, u)$, and using Definition 19, identifies the stratum integral with

$$\int_B \left(\int g_x(a, u)\, d\nu^U_x(u)\right) \pi_x(a)\, da = \int_B m^{\mathrm{F}}_x(a)\pi_x(a)\, da.$$

The observed-margin identity gives

$$\int_{\{X=x,\ A\in B\}} Y\, dP^{\mathrm{F}} = \int_{\{X=x,\ A\in B\}} Y\, dP.$$

The regression-version property of $\mu^P_{x,\mathrm{cont}}$ in Definition 11, together with the conditional-density law, gives

$$\int_{\{X=x,\ A\in B\}} Y\, dP = p_x \int_B \mu^P_{x,\mathrm{cont}}(a)\pi_x(a)\, da.$$

Combining the three displays and using $p_x > 0$,

$$\int_B m^{\mathrm{F}}_x(a)\pi_x(a)\, da = \int_B \mu^P_{x,\mathrm{cont}}(a)\pi_x(a)\, da \qquad (B \subseteq [0, \bar{\delta}]).$$

The two products are integrable on $[0, \bar{\delta}]$: Definitions 11 and 19 and Assumption 8 give continuity of the two mean functions on this compact interval, and Assumption 4 gives the envelope $\pi_x(a) \le c_+ a^\kappa \le c_+$ almost everywhere on $[0, \bar{\delta}]$. Equality of all measurable set integrals therefore gives

$$m^{\mathrm{F}}_x(a)\pi_x(a) = \mu^P_{x,\mathrm{cont}}(a)\pi_x(a) \quad \text{for Lebesgue-a.e. } a \in [0, \bar{\delta}].$$

For Lebesgue-a.e. $a \in (0, \bar{\delta}]$, Assumption 4 also gives

$$\pi_x(a) \ge c_- a^\kappa > 0.$$

Since the endpoint 0 has Lebesgue measure zero, the two continuous functions m^{F}_x and $\mu^P_{x,\mathrm{cont}}$ agree Lebesgue-a.e. on $[0, \bar{\delta}]$. Applying the continuous-version uniqueness principle in Lemma 18,

$$m^{\mathrm{F}}_x(a) = \mu^P_{x,\mathrm{cont}}(a) \qquad \text{for every } a \in [0, \bar{\delta}].$$

2. Fix $\delta \in [0, \bar{\delta}]$. Since $\bar{\delta} \le 1$, also $\delta \in [0,1]$. On the full-data carrier define

$$\varphi_\delta^{\mathrm{ret}}(\omega) = \begin{cases} Y(\omega), & \delta < A(\omega), \\ 0, & \delta \ge A(\omega), \end{cases} \qquad \varphi_\delta^{\mathrm{atom}}(\omega) = \begin{cases} Y(\delta)(\omega), & A(\omega) \le \delta, \\ 0, & A(\omega) > \delta. \end{cases}$$

The common consistency null set in Assumption 6, the observed-treatment support $A \in [0,1]$ supplied by the observed-margin component of Definition 27 and transferred to the full-data law, and Definition 6 give

$$Y(d_\delta(A)) = \varphi_\delta^{\mathrm{ret}} + \varphi_\delta^{\mathrm{atom}} \quad P^{\mathrm{F}}\text{-a.s.}$$

On the branch $\delta < A$, treatment support puts A in $[0,1]$, the clamp policy returns A, and consistency identifies $Y(A)$ with the observed outcome Y. On the branch $A \le \delta$, the clamp policy returns δ, so the displayed decomposition is the corresponding algebraic split. The retained summand is measurable because it is built from the observed outcome and the measurable threshold set $\{A > \delta\}$. The atom summand is measurable because the potential-outcome process in Definition 27 is jointly measurable and $\{A \le \delta\}$ is measurable. For integrability, the observed outcome support bounds $\varphi_\delta^{\mathrm{ret}}$ by one almost surely, and Assumption 6 identifies $Y(\delta)$ with the $[0,1]$-valued structural response $g_X(\delta, U)$ on the consistency event; hence $\varphi_\delta^{\mathrm{atom}}$ is also bounded by one almost surely.

3. The retained summand depends only on the observed coordinates. The observed-margin identity therefore gives

$$\int \varphi_\delta^{\mathrm{ret}}\, dP^{\mathrm{F}} = \int Y\, \mathbf{1}\{A > \delta\}\, dP = \nu_\delta(P),$$

where $\nu_\delta(P)$ is defined in Definition 7.

4. For the atom summand, decompose over strata and use the conditional product law for $p_x^{-1} P^{\mathrm{F}}|_{\{X=x\}}$, now with the dose argument fixed at δ and the treatment event $A \le \delta$:

$$\int \varphi_\delta^{\mathrm{atom}}\, dP^{\mathrm{F}} = \sum_{x \in \mathcal{X}} p_x \left(\int_{[0,\delta]} \pi_x(a)\, da \right) m_x^{\mathrm{F}}(\delta).$$

By Definition 7,

$$q_x(\delta) = \int_{[0,\delta]} \pi_x(a)\, da,$$

so

$$\int \varphi_\delta^{\mathrm{atom}}\, dP^{\mathrm{F}} = \sum_{x \in \mathcal{X}} p_x q_x(\delta) m_x^{\mathrm{F}}(\delta).$$

5. Combining the pathwise decomposition, integrability, and the two integral identities for the retained and atom summands, and using Definition 21,

$$\psi_\delta(P^{\mathrm{F}}) = \nu_\delta(P) + \sum_{x \in \mathcal{X}} p_x q_x(\delta) m_x^{\mathrm{F}}(\delta).$$

Applying the pointwise identity $m_x^{\mathrm{F}} = \mu_{x,\mathrm{cont}}^P$ on $[0, \bar{\delta}]$ at $a = \delta$ gives

$$\psi_\delta(P^{\mathrm{F}}) = \nu_\delta(P) + \sum_{x \in \mathcal{X}} p_x q_x(\delta) \mu_{x,\mathrm{cont}}^P(\delta).$$

The right-hand side is exactly $\theta_{\delta,\mathrm{cont}}(P)$ by Definition 25. Since $\delta \in [0, \bar{\delta}]$ was arbitrary, the asserted equality holds throughout the declared threshold range.

□

Proof of Proposition 5. 1. The continuity-regime hypotheses give

$$J > 0, \quad 0 \le \kappa, \quad 0 < c_-, \quad c_- \le \kappa + 1 \le c_+, \quad 0 < p_{\min} \le \frac{1}{J}, \quad 0 < \bar{\delta} < 1,$$

and also $0 < \alpha < 1/2$. The displayed inequalities are the continuity design constants needed in Proposition 4. The strict inequality $0 < p_{\min}$, together with Assumption 3, gives positive mass to every stratum of any $P \in \mathcal{M}_{\mathrm{cont}}$.

2. Fix $P \in \mathcal{M}_{\mathrm{cont}}$. Put

$$\mathsf{d}_{\mathrm{obs}}(o) = (X(o), A(o)), \qquad \nu = P \circ \mathsf{d}_{\mathrm{obs}}^{-1},$$

and choose a measurable conditional-mean version $\mu_x : \mathbb{R} \to \mathbb{R}$ whose restriction to $[0, \bar{\delta}]$ is the continuous extension $\mu^P_{x,\mathrm{cont}}$ in Definition 11. Take a Borel regular conditional response kernel

$$\mathsf{K}^0_{x,a}(\cdot) = \Pr_P\{Y \in \cdot \mid X = x, A = a\}$$

for Y given (X, A). Equivalently, for Borel $H \subseteq \mathcal{X} \times \mathbb{R}$ and $C \subseteq \mathbb{R}$,

$$P\{\mathsf{d}_{\mathrm{obs}}(O) \in H,\ Y \in C\} = \int_H \mathsf{K}^0_{x,a}(C)\, \nu(d(x,a)).$$

Let

$$\mathrm{clip}(y) = \min\{\max\{y, 0\}, 1\}, \qquad \mathsf{K}^1_{x,a} = (\mathrm{clip})_\# \mathsf{K}^0_{x,a}, \qquad \rho_P(x,a) = \int y\, \mathsf{K}^1_{x,a}(dy).$$

Because $Y \in [0,1]$ almost surely, the raw conditional kernel K^0 and its clipped pushforward K^1 agree for ν-almost every (x,a), so their conditional means agree there. The conditional-distribution formula for conditional expectation, combined with the defining conditional-expectation version in Definition 24, gives directly

$$\rho_P(X, A) = \mu_X(A) \qquad P\text{-almost surely.}$$

Define the total measurable response mean

$$\eta_P(x,a) = \begin{cases} \mathrm{clip}\{\mu_x(a)\}, & a \in [0, \bar{\delta}], \\ \rho_P(x,a), & a \notin [0, \bar{\delta}]. \end{cases}$$

Then $\eta_P(x,a) \in [0,1]$ for every (x,a), and $a \mapsto \eta_P(x,a)$ is continuous on $[0, \bar{\delta}]$ for every x. Moreover,

$$\eta_P(X, A) = \rho_P(X, A) = \mathbb{E}_P[Y \mid X, A] \qquad P\text{-almost surely.}$$

On the strip, the displayed almost-sure equality and $\rho_P \in [0,1]$ show that clipping $\mu_X(A)$ leaves the conditional mean unchanged. Off the strip, the equality is the definition of η_P.

3. Let
$$\mathcal{Z}_\eta = \left\{(x,a) : \int y \, \mathsf{K}^1_{x,a}(dy) \neq \eta_P(x,a)\right\}$$
and define
$$\mathsf{K}^\eta_{x,a} = \begin{cases} \delta_{\eta_P(x,a)}, & (x,a) \in \mathcal{Z}_\eta, \\ \mathsf{K}^1_{x,a}, & (x,a) \notin \mathcal{Z}_\eta. \end{cases}$$
The almost-sure agreement of η_P with the outcome mean of K^1 gives $\nu(\mathcal{Z}_\eta) = 0$, so replacing K^1 by K^η preserves the joint observed law. Also
$$\int y \, \mathsf{K}^\eta_{x,a}(dy) = \eta_P(x,a), \qquad \mathsf{K}^\eta_{x,a}([0,1]) = 1$$
for every (x,a). With λ denoting Lebesgue measure on $[0,1]$, define
$$\mathsf{Q}_x(a,u) = \sup\left\{s \in [0,1] : \mathsf{K}^\eta_{x,a}([0,s]) < u\right\}, \qquad u \in [0,1].$$
The strict-superlevel sets of this supremum are countable unions over rational thresholds, so $(x,a,u) \mapsto \mathsf{Q}_x(a,u)$ is jointly Borel measurable. The generalized-inverse identity on left rays gives
$$(\mathsf{Q}_x(a,\cdot))_\# \lambda = \mathsf{K}^\eta_{x,a}.$$
Construct P^{F} on the latent carrier $\mathcal{U} = [0,1]$ by first sampling $(x,a,u) \sim \nu \otimes \lambda$, taking the latent variable to be $U = u$, and setting
$$X = x, \qquad A = a, \qquad Y = \mathsf{Q}_x(a,u).$$
Define the structural response maps and potential outcomes by
$$g_x(b,u) = \mathsf{Q}_x(b,u), \qquad Y(b) = g_X(b,U) = \mathsf{Q}_X(b,U), \qquad b \in [0,1].$$
The simultaneous definition gives Assumption 6. To verify Assumption 7, it suffices in this finite-stratum setting to check the bounded-Borel product-moment factorization used by the construction. For every stratum x, every bounded Borel $\varphi : \mathbb{R} \to \mathbb{R}$, and every bounded Borel $\chi : [0,1] \to \mathbb{R}$, let $S_x = \{(x',a) : x' = x\}$. The product law $\nu \otimes \lambda$ gives
$$\int_{\{X=x\}} \varphi(A)\chi(U)\, dP^{\mathrm{F}} = \left(\int_{S_x} \varphi(a)\, \nu(d(x',a))\right)\left(\int_0^1 \chi(u)\, d\lambda(u)\right),$$
$$\int_{\{X=x\}} \varphi(A)\, dP^{\mathrm{F}} = \int_{S_x} \varphi(a)\, \nu(d(x',a)), \qquad \int_{\{X=x\}} \chi(U)\, dP^{\mathrm{F}} = \nu(S_x) \int_0^1 \chi(u)\, d\lambda(u),$$
and $P^{\mathrm{F}}(X=x) = \nu(S_x)$. Substitution yields
$$P^{\mathrm{F}}(X=x) \int_{\{X=x\}} \varphi(A)\chi(U)\, dP^{\mathrm{F}} = \left(\int_{\{X=x\}} \varphi(A)\, dP^{\mathrm{F}}\right)\left(\int_{\{X=x\}} \chi(U)\, dP^{\mathrm{F}}\right).$$
The observed margin is P: the conditional law of Y given $(X,A) = (x,a)$ is $\mathsf{K}^\eta_{x,a}$, this kernel equals K^1 for ν-almost every (x,a), and K^1 agrees ν-almost surely with the original

conditional response law because $Y \in [0,1]$ P-almost surely. Finally, for every stratum x and every $a \in [0, \bar{\delta}]$, Definition 19 and the quantile identity give

$$m_x^{\mathrm{F}}(a) = P^{\mathrm{F}}(X = x)^{-1} \int_{\{X=x\}} \mathsf{Q}_x(a, U)\, dP^{\mathrm{F}} = \int_0^1 \mathsf{Q}_x(a, u)\, d\lambda(u) = \eta_P(x, a),$$

where $P^{\mathrm{F}}(X = x) > 0$ follows from $0 < p_{\min}$ and Assumption 3. Since $a \mapsto \eta_P(x, a)$ is continuous on $[0, \bar{\delta}]$, Assumption 8 holds. Therefore

$$P^{\mathrm{F}} \in \mathcal{M}^{\mathrm{F}}_{\mathrm{cont}}, \qquad P^{\mathrm{F}}_{\mathrm{obs}} = P,$$

with the observed margin understood as in Definition 29. This proves surjectivity.

4. Now fix $n \in \mathbb{N}$ and $\delta \in [0, \bar{\delta}]$. For every $P^{\mathrm{F}} \in \mathcal{M}^{\mathrm{F}}_{\mathrm{cont}}$, the design constants recorded above allow Proposition 4 to identify the causal and observed continuity targets:

$$\psi_\delta(P^{\mathrm{F}}) = \theta_{\delta,\mathrm{cont}}(P^{\mathrm{F}}_{\mathrm{obs}}).$$

The observed product sample law in Definition 2 depends on P^{F} only through the observed margin $P^{\mathrm{F}}_{\mathrm{obs}}$. Hence, for every observed-sample estimator $\widehat{\theta}$ satisfying the measurability and range requirements in Definition 28,

$$\left\{ \int \left| \widehat{\theta}(z) - \psi_\delta(P^{\mathrm{F}}) \right| d(P^{\mathrm{F}}_{\mathrm{obs}})^{\otimes n}(z) : P^{\mathrm{F}} \in \mathcal{M}^{\mathrm{F}}_{\mathrm{cont}} \right\} = \left\{ \int \left| \widehat{\theta}(z) - \theta_{\delta,\mathrm{cont}}(P) \right| dP^{\otimes n}(z) : P \in \mathcal{M}_{\mathrm{cont}} \right\}.$$

The inclusion from left to right uses $P^{\mathrm{F}}_{\mathrm{obs}} \in \mathcal{M}_{\mathrm{cont}}$; the inclusion from right to left uses surjectivity of the observed-margin map Γ onto $\mathcal{M}_{\mathrm{cont}}$. Taking the supremum of equal value sets and then the same infimum over observed-sample estimators in Definition 28 yields

$$R^\star_{n,\mathrm{cont},\mathrm{F}} = R^\star_{n,\mathrm{cont}}.$$

The argument also covers $n = 0$, since it uses only equality of the defining value sets.

5. The interval criteria are handled in the same way. For every observed-sample interval procedure C_n satisfying the measurability requirement in Definition 28, the target identity and surjectivity give the equivalence

$$\forall P^{\mathrm{F}} \in \mathcal{M}^{\mathrm{F}}_{\mathrm{cont}}, \quad 1 - \alpha \le (P^{\mathrm{F}}_{\mathrm{obs}})^{\otimes n} \{ z : \psi_\delta(P^{\mathrm{F}}) \in C_n(z) \}$$

if and only if

$$\forall P \in \mathcal{M}_{\mathrm{cont}}, \quad 1 - \alpha \le P^{\otimes n} \{ z : \theta_{\delta,\mathrm{cont}}(P) \in C_n(z) \}.$$

Using the pointwise interval length from Definition 15, the same observed-margin argument gives equality of expected-length value sets:

$$\left\{ \int \mathrm{len}\{C_n(z)\}\, d(P^{\mathrm{F}}_{\mathrm{obs}})^{\otimes n}(z) : P^{\mathrm{F}} \in \mathcal{M}^{\mathrm{F}}_{\mathrm{cont}} \right\} = \left\{ \int \mathrm{len}\{C_n(z)\}\, dP^{\otimes n}(z) : P \in \mathcal{M}_{\mathrm{cont}} \right\}.$$

Taking the supremum over laws and then the infimum over honest interval procedures in Definition 28 yields

$$L^\star_{n,\mathrm{cont},\mathrm{F}} = L^\star_{n,\mathrm{cont}}.$$

Together with the risk equality, this proves the two displayed identities for every $n \in \mathbb{N}$ and every $\delta \in [0, \bar{\delta}]$.

□

Proof of Theorem 5. 1. Lower experiments for the observed criteria. For $n \geq 1$ and a threshold sequence write

$$r_{n,\mathrm{root}} = (n : \mathbb{R})^{-1/2}, \qquad a_{n,\mathrm{atom}} = \delta_n^{\kappa+1}, \qquad s_n = r_{n,\mathrm{root}} + a_{n,\mathrm{atom}},$$

as in Definition 26. The lower-bound witnesses all use the same design:

$$\Pr(X = x) = \frac{1}{J}, \qquad \pi_x(a) = (\kappa+1)a^\kappa, \quad 0 \leq a \leq 1.$$

Since $0 < J$, $0 \leq \kappa$, $0 < p_{\min} \leq 1/J$, and $c_- \leq \kappa+1 \leq c_+$, this design satisfies the stratum-mass and thinning requirements in Definition 24. For the root term, set

$$Q_{\mathrm{root},0}(Y = 1 \mid X, A) = \frac{1}{2}, \qquad Q_{\mathrm{root},1}(Y = 1 \mid X, A) = \frac{1}{2} + \varepsilon_n, \qquad \varepsilon_n = \frac{1}{4} r_{n,\mathrm{root}}.$$

The two continuous regressions lie in $[0,1]$, their targets differ by

$$|\theta_{\delta_n,\mathrm{cont}}(Q_{\mathrm{root},1}) - \theta_{\delta_n,\mathrm{cont}}(Q_{\mathrm{root},0})| = \varepsilon_n,$$

and the product chi-square divergence obeys

$$1 + \chi^2(Q_{\mathrm{root},1}^{\otimes n}, Q_{\mathrm{root},0}^{\otimes n}) \leq \exp\{4n\varepsilon_n^2\} = \exp(1/4).$$

For the atom term, let $\varphi : \mathbb{R} \to [0,1]$ be the fixed smooth treatment bump with $\varphi(0) = 1$ and $\varphi(u) = 0$ whenever $|u| \geq 1$. Put $a_\star = 1/4$ and choose

$$A_\star = 8(\kappa+1)a_\star^2(\bar{\delta}+1)^\kappa, \qquad K_{\mathrm{risk}} = 1 + A_\star, \qquad h_{n,\mathrm{risk}} = \frac{1}{K_{\mathrm{risk}} n}.$$

With

$$q_{n,\mathrm{risk}}(a) = a_\star\, \varphi\left(\frac{a - \delta_n}{h_{n,\mathrm{risk}}}\right),$$

define $Q_{\mathrm{atom},0}$ by regression $1/2$ and $Q_{\mathrm{atom},1}$ by regression $1/2 + q_{n,\mathrm{risk}}(a)$. The bump properties give a continuous perturbation with $0 \leq q_{n,\mathrm{risk}} \leq a_\star$, support contained in $[\delta_n - h_{n,\mathrm{risk}}, \delta_n + h_{n,\mathrm{risk}}]$, and value $q_{n,\mathrm{risk}}(\delta_n) = a_\star$. Therefore the atom contribution at the clamp point gives

$$a_\star\, \delta_n^{\kappa+1} \leq |\theta_{\delta_n,\mathrm{cont}}(Q_{\mathrm{atom},1}) - \theta_{\delta_n,\mathrm{cont}}(Q_{\mathrm{atom},0})|\,.$$

The same localization and boundedness give the design-integral estimate behind the Bernoulli chi-square calculation,

$$\mathbb{E}\big[4q_{n,\mathrm{risk}}(A)^2\big] \leq 8(\kappa+1)a_\star^2 h_{n,\mathrm{risk}}(\delta_n + h_{n,\mathrm{risk}})^\kappa,$$

under the displayed common design. Since $h_{n,\mathrm{risk}} \leq 1$ and $\delta_n \leq \bar{\delta}$, the exponent is at most $A_\star/K_{\mathrm{risk}} \leq 1$, and hence

$$1 + \chi^2(Q_{\mathrm{atom},1}^{\otimes n}, Q_{\mathrm{atom},0}^{\otimes n}) \leq \exp\big\{n\, 8(\kappa+1)a_\star^2 h_{n,\mathrm{risk}}(\delta_n + h_{n,\mathrm{risk}})^\kappa\big\} \leq e.$$

We use the following two-point risk bound. For probability laws $\mathfrak{Q}_0, \mathfrak{Q}_1$ on the sample space and real target values ϑ_0, ϑ_1, every estimator $\mathfrak{t}$ satisfies

$$\max_{\iota \in \{0,1\}} \mathbb{E}_{\mathfrak{Q}_\iota} |\mathfrak{t} - \vartheta_\iota| \geq \frac{|\vartheta_1 - \vartheta_0|}{2}\{1 - \mathrm{TV}(\mathfrak{Q}_0, \mathfrak{Q}_1)\}.$$

To verify this, fix a probability measure Λ dominating both laws and define the common-part measure

$$\mathfrak{Q}_\wedge(E) = \int_E \min\left\{\frac{\mathrm{d}\mathfrak{Q}_0}{\mathrm{d}\Lambda}, \frac{\mathrm{d}\mathfrak{Q}_1}{\mathrm{d}\Lambda}\right\} \mathrm{d}\Lambda.$$

Its total mass is $1 - \mathrm{TV}(\mathfrak{Q}_0, \mathfrak{Q}_1)$, and integration of $|\mathfrak{t} - \vartheta_0| + |\mathfrak{t} - \vartheta_1| \geq |\vartheta_1 - \vartheta_0|$ against $\mathfrak{Q}_\wedge$ gives the display. When $\mathfrak{Q}_1 \ll \mathfrak{Q}_0$, Cauchy–Schwarz applied to the likelihood ratio gives

$$\mathrm{TV}(\mathfrak{Q}_0, \mathfrak{Q}_1) \leq \frac{1}{2}\sqrt{\chi^2(\mathfrak{Q}_1, \mathfrak{Q}_0)}.$$

For the root pair, the product bound gives $\chi^2 \leq e^{1/4} - 1 < 4$; for the atom pair it gives $\chi^2 \leq e - 1 < 4$. Hence the factors $1 - \frac{1}{2}\sqrt{\chi^2}$ are positive in both two-point applications. The two target separations therefore yield constants $c_{\mathrm{R},0}, c_{\mathrm{R},1} > 0$ such that, eventually,

$$c_{\mathrm{R},0} r_{n,\mathrm{root}} \leq R^\star_{n,\mathrm{cont}}, \qquad c_{\mathrm{R},1} a_{n,\mathrm{atom}} \leq R^\star_{n,\mathrm{cont}}.$$

Taking $c_{\mathrm{R}} = \min\{c_{\mathrm{R},0}, c_{\mathrm{R},1}\}/2$ gives

$$c_{\mathrm{R}} s_n \leq R^\star_{n,\mathrm{cont}}.$$

For honest length, set

$$\eta_\alpha = 1 - 2\alpha, \qquad \chi_\alpha = \eta_\alpha^2/16, \qquad \tau_\alpha = \eta_\alpha/16.$$

For the root length witnesses, let $Q_{\mathrm{len,root},0}$ and $Q_{\mathrm{len,root},1}$ be the global-shift pair with $\varepsilon_{n,\mathrm{len}} = \tau_\alpha r_{n,\mathrm{root}}$. Then

$$1 + \chi^2(Q^{\otimes n}_{\mathrm{len,root},1}, Q^{\otimes n}_{\mathrm{len,root},0}) \leq \exp(\eta_\alpha^2/64), \qquad \chi^2(Q^{\otimes n}_{\mathrm{len,root},1}, Q^{\otimes n}_{\mathrm{len,root},0}) \leq \chi_\alpha.$$

The second inequality follows from $0 < \eta_\alpha < 1$, $e^x \leq 1 + x + x^2$ at $x = \eta_\alpha^2/64$, and $x + x^2 \leq \eta_\alpha^2/16$. For the atom length witnesses, let $Q_{\mathrm{len,atom},0}$ and $Q_{\mathrm{len,atom},1}$ be the localized-bump pair with the same amplitude $a_\star$ and choose

$$K_{\mathrm{len}} = \eta_\alpha^2 + 64A_\star, \qquad h_{n,\mathrm{len}} = \frac{\eta_\alpha^2}{K_{\mathrm{len}} n}.$$

The localized-bump exponent satisfies

$$n\, 8(\kappa + 1) a_\star^2 h_{n,\mathrm{len}} (\delta_n + h_{n,\mathrm{len}})^\kappa \leq \frac{A_\star \eta_\alpha^2}{K_{\mathrm{len}}} \leq \frac{\eta_\alpha^2}{64},$$

so the same exponential-to-polynomial calculation gives

$$\chi^2(Q^{\otimes n}_{\mathrm{len,atom},1}, Q^{\otimes n}_{\mathrm{len,atom},0}) \leq \chi_\alpha.$$

For either length-witness pair, write the first law as $\mathfrak{Q}_0$, the second as $\mathfrak{Q}_1$, and the two target values as ϑ_0, ϑ_1. If an interval covers ϑ_ι under $\mathfrak{Q}_\iota$ with probability at least $1 - \alpha$ for $\iota = 0, 1$, then total variation and the union bound give

$$\mathfrak{Q}_0\{\vartheta_0, \vartheta_1 \in C_n\} \geq 1 - 2\alpha - \mathrm{TV}(\mathfrak{Q}_0, \mathfrak{Q}_1).$$

Together with $\mathrm{TV}(\mathfrak{Q}_0,\mathfrak{Q}_1)\le \frac{1}{2}\sqrt{\chi^2(\mathfrak{Q}_1,\mathfrak{Q}_0)}$, the displayed chi-square bounds give

$$1-2\alpha-\frac{1}{2}\sqrt{\chi^2(\mathfrak{Q}_1,\mathfrak{Q}_0)}\ge \eta_\alpha-\frac{1}{2}\sqrt{\chi_\alpha}=\frac{7}{8}\eta_\alpha>0.$$

On the event $\{\vartheta_0,\vartheta_1\in C_n\}$, the interval length is at least $|\vartheta_1-\vartheta_0|$. Thus there are $c_{\mathrm{L},0},c_{\mathrm{L},1}>0$ with

$$c_{\mathrm{L},0}r_{n,\mathrm{root}}\le L^\star_{n,\mathrm{cont}},\qquad c_{\mathrm{L},1}a_{n,\mathrm{atom}}\le L^\star_{n,\mathrm{cont}}$$

in the extended nonnegative-real order, where nonnegative real lower bounds are embedded as finite extended values. With $c_{\mathrm{L}}=\min\{c_{\mathrm{L},0},c_{\mathrm{L},1}\}/2$, the combination is by the case split $c_{\mathrm{L},0}r_{n,\mathrm{root}}\le c_{\mathrm{L},1}a_{n,\mathrm{atom}}$ or its complement, yielding

$$c_{\mathrm{L}}s_n\le L^\star_{n,\mathrm{cont}}.$$

2. Upper risk for the fixed fallback estimator. Fix $P\in\mathcal{M}_{\mathrm{cont}}$, a sample split $B_n=(I_0,I_1,I_2)$, and $\delta=\delta_n$. Define

$$\widehat{\nu}_n=|I_0|^{-1}\sum_{i\in I_0}Y_i\mathbf{1}\{A_i>\delta\},\qquad \widehat{b}_{n,x}=|I_1|^{-1}\sum_{i\in I_1}\mathbf{1}\{X_i=x,\ A_i\le\delta\},$$

$$\nu_\delta(P)=\mathbb{E}_P[Y\mathbf{1}\{A>\delta\}],\qquad b_x(P,\delta)=p_xq_x(\delta),\qquad \mathfrak{e}_\delta=\frac{c_+\delta^{\kappa+1}}{\kappa+1}.$$

The estimator in Algorithm 3 is

$$\widehat{\theta}_{n,\mathrm{cont}}=\Pi_{[0,1]}\left(\widehat{\nu}_n+\frac{1}{2}\sum_x\widehat{b}_{n,x}\right),$$

where $\Pi_{[0,1]}(t)=\min\{1,\max\{0,t\}\}$. By Definition 25, the target atom contribution is

$$A_{\delta,\mathrm{cont}}(P)=\sum_x b_x(P,\delta)\mu^P_{x,\mathrm{cont}}(\delta),\qquad \theta_{\delta,\mathrm{cont}}(P)=\nu_\delta(P)+A_{\delta,\mathrm{cont}}(P).$$

Since $\theta_{\delta,\mathrm{cont}}(P)\in[0,1]$, projection is contractive toward the target. Adding and subtracting the fallback atom term gives

$$\begin{aligned}\left|\widehat{\theta}_{n,\mathrm{cont}}-\theta_{\delta,\mathrm{cont}}(P)\right|&\le|\widehat{\nu}_n-\nu_\delta(P)|+\left|\frac{1}{2}\sum_x\widehat{b}_{n,x}-\sum_x b_x(P,\delta)\mu^P_{x,\mathrm{cont}}(\delta)\right|\\&\le|\widehat{\nu}_n-\nu_\delta(P)|+\frac{1}{2}\sum_x\left|\widehat{b}_{n,x}-b_x(P,\delta)\right|+\sum_x b_x(P,\delta)\left|\frac{1}{2}-\mu^P_{x,\mathrm{cont}}(\delta)\right|\\&\le|\widehat{\nu}_n-\nu_\delta(P)|+\sum_x\left\{\frac{1}{2}\left|\widehat{b}_{n,x}-b_x(P,\delta)\right|+\frac{1}{2}\mathfrak{e}_\delta\right\},\end{aligned}$$

Here

$$0\le b_x(P,\delta)=p_xq_x(\delta)\le\frac{c_+\delta^{\kappa+1}}{\kappa+1}=\mathfrak{e}_\delta,$$

by the thinning envelope integrated over $[0,\delta]$. The retained and atom indicators are $[0,1]$-valued block averages, so Cauchy–Schwarz and the variance bound $\mathrm{Var}(V)\le 1/4$ for $V\in[0,1]$ give

$$\mathbb{E}_P\left|\widehat{\nu}_n-\nu_\delta(P)\right|\le\frac{1}{2}|I_0|^{-1/2},\qquad \mathbb{E}_P\left|\widehat{b}_{n,x}-b_x(P,\delta)\right|\le\frac{1}{2}|I_1|^{-1/2}.$$

Integrating the pointwise bound therefore gives the explicit estimate

$$\mathbb{E}_{P^{\otimes n}}\left|\widehat{\theta}_{n,\mathrm{cont}}-\theta_{\delta,\mathrm{cont}}(P)\right|\le\frac{1}{2}|I_0|^{-1/2}+J\left\{\frac{1}{4}|I_1|^{-1/2}+\frac{1}{2}\frac{c_+\delta^{\kappa+1}}{\kappa+1}\right\}.$$

For all $n\ge 8$, admissibility of B_n gives $|I_j|^{-1/2}\le 3n^{-1/2}$ for $j=0,1$. Hence, uniformly over P,

$$\sup_{P\in\mathcal{M}_{\mathrm{cont}}}\mathbb{E}_{P^{\otimes n}}\left|\widehat{\theta}_{n,\mathrm{cont}}-\theta_{\delta_n,\mathrm{cont}}(P)\right|\le C_{\mathrm{R}}s_n$$

for a constant $C_{\mathrm{R}}>0$ depending only on the declared constants. The same estimator is an admissible competitor in Definition 28, so

$$R^{\star}_{n,\mathrm{cont}}\le\sup_{P\in\mathcal{M}_{\mathrm{cont}}}\mathbb{E}_{P^{\otimes n}}\left|\widehat{\theta}_{n,\mathrm{cont}}-\theta_{\delta_n,\mathrm{cont}}(P)\right|.$$

3. Coverage and length of the continuity interval. Let

$$t_{0,\alpha}=\sqrt{\frac{\log(4/\alpha)}{2|I_0|}},\qquad t_{1,\alpha}=\sqrt{\frac{\log(4/\alpha)}{2|I_1|}},\qquad \widehat{B}_n=\sum_x\widehat{b}_{n,x}.$$

The interval in Algorithm 4 is

$$\mathrm{CI}_{n,\mathrm{cont}}=\left[\max\{0,\widehat{\theta}_{n,\mathrm{cont}}-(t_{0,\alpha}+t_{1,\alpha}+\widehat{B}_n/2)\},\,\min\{1,\widehat{\theta}_{n,\mathrm{cont}}+(t_{0,\alpha}+t_{1,\alpha}+\widehat{B}_n/2)\}\right].$$

The coverage argument uses the following unit-range specialization of [Hoeffding, 1963, Theorem 2]: an average of m independent $[0,1]$-valued variables has two-sided deviation probability at radius $t>0$ at most $2\exp(-2mt^2)$. The total empirical atom mass is itself one such average, because the strata partition the sample space:

$$\widehat{B}_n=\sum_x\widehat{b}_{n,x}=|I_1|^{-1}\sum_{i\in I_1}\mathbf{1}\{A_i\le\delta\}.$$

Its population mean is

$$B_\delta(P)=\sum_x b_x(P,\delta)=P(A\le\delta).$$

Applied to the retained-course average and to this total lower-threshold atom average, with the radii in Algorithm 4, the Hoeffding bound gives

$$P^{\otimes n}(|\widehat{\nu}_n-\nu_\delta(P)|>t_{0,\alpha})\le\frac{\alpha}{2},\qquad P^{\otimes n}\left(\left|\widehat{B}_n-B_\delta(P)\right|>t_{1,\alpha}\right)\le\frac{\alpha}{2}.$$

On the complement of the two events, set

$$A_\delta(P)=\sum_x b_x(P,\delta)\mu^P_{x,\mathrm{cont}}(\delta).$$

The regression values lie in $[0,1]$, so the unknown clamp atom contribution satisfies

$$0\le A_\delta(P)\le B_\delta(P).$$

Since the concentration event also gives $B_\delta(P) \le \widehat{B}_n + t_{1,\alpha}$, the quantity $\widehat{B}_n/2 - A_\delta(P)$ lies between $-\widehat{B}_n/2 - t_{1,\alpha}$ and $\widehat{B}_n/2$. Hence

$$\left|\frac{1}{2}\widehat{B}_n - A_\delta(P)\right| \le t_{1,\alpha} + \widehat{B}_n/2.$$

Using Definition 25 and the contraction of projection onto $[0,1]$ toward $\theta_{\delta,\mathrm{cont}}(P) \in [0,1]$, this gives

$$\left|\widehat{\theta}_{n,\mathrm{cont}} - \theta_{\delta,\mathrm{cont}}(P)\right| \le t_{0,\alpha} + t_{1,\alpha} + \widehat{B}_n/2,$$

so $\theta_{\delta,\mathrm{cont}}(P) \in \mathrm{CI}_{n,\mathrm{cont}}$. The union bound yields uniform coverage at least $1-\alpha$. Also,

$$\mathrm{len}(\mathrm{CI}_{n,\mathrm{cont}}) \le 2t_{0,\alpha} + 2t_{1,\alpha} + \widehat{B}_n.$$

Moreover, the displayed identity for $\widehat{B}_n$ gives

$$\mathbb{E}_{P^{\otimes n}}\widehat{B}_n = B_\delta(P) = \sum_x b_x(P,\delta) \le J\mathfrak{e}_\delta.$$

Write $[u]_{\overline{\mathbb{R}}_+}$ for the canonical embedding of a nonnegative real u into the extended nonnegative reals. Consequently, before the final uniform simplification one has the extended-real length bound

$$\int^{+} \mathrm{len}(\mathrm{CI}_{n,\mathrm{cont}})\, dP^{\otimes n} \le \left[2\sqrt{\frac{\log(4/\alpha)}{2|I_0|}} + 2\sqrt{\frac{\log(4/\alpha)}{2|I_1|}} + J\frac{c_+\delta^{\kappa+1}}{\kappa+1}\right]_{\overline{\mathbb{R}}_+}.$$

Since $|I_1|^{-1/2} \ge 0$, this immediately implies the deliberately weaker bound

$$\int^{+} \mathrm{len}(\mathrm{CI}_{n,\mathrm{cont}})\, dP^{\otimes n} \le \left[2\sqrt{\frac{\log(4/\alpha)}{2|I_0|}} + 2\sqrt{\frac{\log(4/\alpha)}{2|I_1|}} + J\left\{\frac{1}{2}|I_1|^{-1/2} + \frac{c_+\delta^{\kappa+1}}{\kappa+1}\right\}\right]_{\overline{\mathbb{R}}_+}.$$

The first display follows from the pointwise inequality $\mathrm{len}(\mathrm{CI}_{n,\mathrm{cont}}) \le 2t_{0,\alpha} + 2t_{1,\alpha} + \widehat{B}_n$ and the preceding expectation calculation; the second only enlarges the finite real upper bound by a nonnegative term. With $|I_j|^{-1/2} \le 3n^{-1/2}$ and $K_\alpha = \sqrt{\log(4/\alpha)/2}$, this gives a constant $C_{\mathrm{L}} > 0$ such that

$$\sup_{P\in\mathcal{M}_{\mathrm{cont}}} \mathbb{E}_{P^{\otimes n}}\, \mathrm{len}(\mathrm{CI}_{n,\mathrm{cont}}) \le C_{\mathrm{L}} s_n$$

and the same comparison holds in the extended nonnegative-real order after embedding the real bound as a finite extended value.

4. Transport to the causal continuity class. Let $P^{\mathrm{F}}_{\mathrm{obs}}$ denote the observed margin in Definition 29. Since the estimator and interval are functions only of the observed coordinates, for every nonnegative or integrable observed-sample functional H,

$$\mathbb{E}_{(P^{\mathrm{F}})^{\otimes n}} H(O_1,\dots,O_n) = \mathbb{E}_{(P^{\mathrm{F}}_{\mathrm{obs}})^{\otimes n}} H(O_1,\dots,O_n),$$

and the analogous identity holds for observed-sample events. The causal minimax criteria in Definition 28 are defined with the full-data product law. For the observed-only estimator and interval used here, the preceding pushforward identity gives the corresponding fixed-procedure risk, coverage, and expected length as the same quantities computed under the observed-margin product law. By Proposition 4, every $P^{\mathrm{F}} \in \mathcal{M}^{\mathrm{F}}_{\mathrm{cont}}$ and every $\delta \in [0, \bar{\delta}]$ satisfy

$$\psi_\delta(P^{\mathrm{F}}) = \theta_{\delta,\mathrm{cont}}(P^{\mathrm{F}}_{\mathrm{obs}}).$$

By Proposition 5, every $P \in \mathcal{M}_{\mathrm{cont}}$ is the observed margin of some member of $\mathcal{M}^{\mathrm{F}}_{\mathrm{cont}}$, and the observed and causal criteria in Definition 28 agree exactly:

$$R^\star_{n,\mathrm{cont},\mathrm{F}} = R^\star_{n,\mathrm{cont}}, \qquad L^\star_{n,\mathrm{cont},\mathrm{F}} = L^\star_{n,\mathrm{cont}}.$$

The bridge identity and observed-margin surjectivity also identify the fixed-procedure observed and causal worst-case quantities. For the fallback estimator,

$$\sup_{P^{\mathrm{F}} \in \mathcal{M}^{\mathrm{F}}_{\mathrm{cont}}} \mathbb{E}_{(P^{\mathrm{F}}_{\mathrm{obs}})^{\otimes n}} \left| \widehat{\theta}_{n,\mathrm{cont}} - \psi_{\delta_n}(P^{\mathrm{F}}) \right| = \sup_{P \in \mathcal{M}_{\mathrm{cont}}} \mathbb{E}_{P^{\otimes n}} \left| \widehat{\theta}_{n,\mathrm{cont}} - \theta_{\delta_n,\mathrm{cont}}(P) \right|.$$

For the interval,

$$\inf_{P^{\mathrm{F}} \in \mathcal{M}^{\mathrm{F}}_{\mathrm{cont}}} (P^{\mathrm{F}}_{\mathrm{obs}})^{\otimes n} \{\psi_{\delta_n}(P^{\mathrm{F}}) \in \mathrm{CI}_{n,\mathrm{cont}}\} = \inf_{P \in \mathcal{M}_{\mathrm{cont}}} P^{\otimes n} \{\theta_{\delta_n,\mathrm{cont}}(P) \in \mathrm{CI}_{n,\mathrm{cont}}\},$$

and the expected lengths satisfy

$$\sup_{P^{\mathrm{F}} \in \mathcal{M}^{\mathrm{F}}_{\mathrm{cont}}} \mathbb{E}_{(P^{\mathrm{F}}_{\mathrm{obs}})^{\otimes n}} \operatorname{len}(\mathrm{CI}_{n,\mathrm{cont}}) = \sup_{P \in \mathcal{M}_{\mathrm{cont}}} \mathbb{E}_{P^{\otimes n}} \operatorname{len}(\mathrm{CI}_{n,\mathrm{cont}}).$$

5. Choice of constants and eventual frontier bounds. Let $c_{\mathrm{R}}, c_{\mathrm{L}}, C_{\mathrm{R}}, C_{\mathrm{L}}$ be the constants obtained above and define

$$c = \min\{c_{\mathrm{R}}, c_{\mathrm{L}}\}, \qquad C = \max\{C_{\mathrm{R}}, C_{\mathrm{L}}\} + c + 1.$$

Then $0 < c < C$. Intersecting the eventual sets from the lower-risk, lower-length, upper-risk, and upper-length bounds with the sample-size threshold $n \geq 8$ gives, for every admissible threshold sequence and every admissible split sequence, all sufficiently large n satisfying; at those sample sizes the positivity of $|I_0|$ and $|I_1|$ used in the coverage and upper-bound arguments follows from split admissibility.

$$c s_n \leq R^\star_{n,\mathrm{cont}} \leq \sup_{P \in \mathcal{M}_{\mathrm{cont}}} \mathbb{E}_{P^{\otimes n}} \left| \widehat{\theta}_{n,\mathrm{cont}} - \theta_{\delta_n,\mathrm{cont}}(P) \right| \leq C s_n,$$

the asserted observed coverage and length upper bounds, and, in the extended nonnegative-real order,

$$c s_n \leq L^\star_{n,\mathrm{cont}}.$$

The equalities and causal bounds follow from the transport identities above with the same c, C, estimator, and interval.

6. Elbow, zero threshold, and fixed positive thresholds. Since $\kappa + 1 > 0$, for $n > 0$,

$$\left(n^{-1/(2(\kappa+1))}\right)^{\kappa+1} = n^{-1/2},$$

and therefore

$$s_n\left(\delta_n = n^{-1/(2(\kappa+1))}\right) = 2n^{-1/2} \asymp n^{-1/2}.$$

At $\delta_n = 0$, the identity $0^{\kappa+1} = 0$ gives

$$s_n(0) = n^{-1/2} \asymp n^{-1/2}.$$

If $\delta_n \to \delta_0 \in (0, \bar{\delta}]$, then continuity of $u \mapsto u^{\kappa+1}$ at the positive point δ_0 gives

$$\delta_n^{\kappa+1} \to \delta_0^{\kappa+1} > 0.$$

The fixed-positive-threshold argument applies the atom-scale lower constructions directly. For the risk criterion, there is a constant $c_{\mathrm{R},+} > 0$ such that eventually

$$c_{\mathrm{R},+}\delta_n^{\kappa+1} \le R^\star_{n,\mathrm{cont}}.$$

If $R^\star_{n,\mathrm{cont}} \to 0$, passing to limits in this eventual inequality would give $c_{\mathrm{R},+}\delta_0^{\kappa+1} \le 0$, contradicting $c_{\mathrm{R},+}\delta_0^{\kappa+1} > 0$. For honest length, the corresponding lower construction gives a constant $c_{\mathrm{L},+} > 0$ with

$$c_{\mathrm{L},+}\delta_n^{\kappa+1} \le L^\star_{n,\mathrm{cont}}$$

eventually in the extended nonnegative-real order. Convergence of $L^\star_{n,\mathrm{cont}}$ to zero would imply $c_{\mathrm{L},+}\delta_0^{\kappa+1} \le 0$, again impossible. Hence neither fixed-positive-threshold criterion tends to zero. □

Proof of Proposition 6. Fix x and δ. By Definition 5, the conditional treatment law in stratum x is represented on $[0, 1]$ by the density π_x. By Definition 6, the clamp map is measurable and satisfies

$$d_\delta(a) = \delta \quad \text{for } a \le \delta, \qquad d_\delta(a) = a \quad \text{for } a > \delta.$$

Splitting the conditional treatment law at $(-\infty, \delta]$ and its complement gives

$$P\{d_\delta(A) \in \cdot \mid X = x\} = P\{A \in \cdot,\ A > \delta \mid X = x\} + P\{A \le \delta \mid X = x\}\, \delta_\delta.$$

Because this conditional treatment law is carried by $[0, 1]$, the retained part is

$$P\{A \in \cdot,\ A > \delta \mid X = x\} = P\{A \in \cdot,\ A \in (\delta, 1] \mid X = x\}.$$

The same density representation, together with $0 \le \delta \le \bar{\delta} < 1$, gives

$$P\{A \le \delta \mid X = x\} = \int_{[0,\delta]} \pi_x(a)\, da = q_x(\delta),$$

where the last equality is the definition of $q_x(\delta)$ in Definition 7. These identities give the asserted pushforward decomposition.

It remains to bound $q_x(\delta)$. The conditional-density and polynomial-thinning conditions in Definition 5 give integrability of π_x on $[0, \delta]$ and, for Lebesgue-a.e. $a \in [0, \delta]$,

$$c_- a^\kappa \le \pi_x(a) \le c_+ a^\kappa.$$

Since $\kappa \ge 0$, the function $a \mapsto a^\kappa$ is integrable on $[0, \delta]$. Integrating the almost-everywhere inequalities yields

$$c_- \int_0^\delta a^\kappa\, da \le q_x(\delta) \le c_+ \int_0^\delta a^\kappa\, da.$$

The regime condition $\kappa \ge 0$ implies $-1 < \kappa$, and with $0 \le \delta$,

$$\int_0^\delta a^\kappa \, da = \frac{\delta^{\kappa+1}}{\kappa+1}.$$

Substitution gives

$$\frac{c_- \, \delta^{\kappa+1}}{\kappa+1} \le q_x(\delta) \le \frac{c_+ \, \delta^{\kappa+1}}{\kappa+1},$$

as required. $\square$

Proof of Proposition 7. Fix the stratum x, and let

$$\mu_0 = \mathrm{Leb} \restriction [0,1], \qquad \mu_x = \mu_0 \text{ with density } \pi_x.$$

Thus μ_x is the conditional natural-treatment measure in stratum x. The endpoints are μ_0-null, so

$$\mu_0\text{-almost every } a \text{ lies in } (0,1).$$

By Assumption 4, for μ_0-almost every $a \in (0,1)$,

$$\pi_x(a) \ge c_- a^\kappa > 0,$$

because $c_- > 0$, $\kappa \ge 0$, and $a^\kappa > 0$ for $a > 0$. Hence $\pi_x > 0$ μ_0-almost everywhere. Since μ_x is obtained from μ_0 by this density, $\mu_x \ll \mu_0$, and therefore

$$\mu_x\text{-almost every } a \text{ lies in } (0,1).$$

Fix such an a, and put

$$y = d_\delta(a) = \max\{a, \delta\},$$

using Definition 6. Since $0 < a < 1$, $0 \le \delta \le \bar{\delta}$, and $\bar{\delta} < 1$, we have

$$0 < y < 1.$$

Let $U \subset \mathbb{R}$ be open with $y \in U$. Choose $\varepsilon > 0$ such that $B(y, \varepsilon) \subset U$, and set

$$r = \min\{\varepsilon, y, 1 - y\}.$$

Then $r > 0$ and

$$B(y, r) \subset U \cap [0,1].$$

Consequently,

$$0 < \mathrm{Leb}(B(y,r)) \le \mu_0(U).$$

It remains to transfer this positivity from μ_0 to μ_x. For a measure with density, the zero-set identity gives, for every open E,

$$\mu_x(E) = 0 \iff \mu_0\big(E \cap \{z : \pi_x(z) > 0\}\big) = 0.$$

If $\mu_x(U) = 0$, this identity gives

$$\mu_0\big(U \cap \{z : \pi_x(z) > 0\}\big) = 0.$$

Because $\pi_x > 0$ μ_0-almost everywhere, the set in this display agrees with U up to a μ_0-null set, contradicting $0 < \mu_0(U)$. Thus $0 < \mu_x(U)$. Since U was arbitrary, $d_\delta(a)$ is a support point of the conditional natural-treatment measure in stratum x. The choice of a was from a full μ_x-measure set, which gives the claimed almost-everywhere support preservation. $\square$

References

Timothy B. Armstrong and Michal Kolesár. Optimal inference in a class of regression models. *Econometrica*, 86(2):655–683, 2018. doi: 10.3982/ECTA14434.

Timothy B. Armstrong and Michal Kolesár. Simple and honest confidence intervals in nonparametric regression. *Quantitative Economics*, 11(1):1–39, 2016. doi: 10.3982/QE1195.

Matteo Bonvini and Edward H. Kennedy. Fast convergence rates for dose-response estimation. *Electronic Journal of Statistics*, 20(2):3839–3907, 2026. doi: 10.1214/26-EJS2573.

T. Tony Cai and Mark G. Low. An adaptation theory for nonparametric confidence intervals. *The Annals of Statistics*, 32(5):1805–1840, 2004a. doi: 10.1214/009053604000000049.

T. Tony Cai and Mark G. Low. An adaptation theory for nonparametric confidence intervals. *The Annals of Statistics*, 32(5):1805–1840, 2004b. doi: 10.1214/009053604000000049.

David L. Donoho and Richard C. Liu. Geometrizing rates of convergence, ii. *The Annals of Statistics*, 19(2):633–667, 1991a. doi: 10.1214/aos/1176348114.

David L. Donoho and Richard C. Liu. Geometrizing rates of convergence, ii. *The Annals of Statistics*, 19(2):633–667, 1991b. doi: 10.1214/aos/1176348114.

Iván Díaz, Nicholas Williams, Katherine L. Hoffman, and Edward J. Schenck. Nonparametric causal effects based on longitudinal modified treatment policies. *Journal of the American Statistical Association*, 118(542):846–857, 2021. doi: 10.1080/01621459.2021.1955691.

Iván Díaz, Nicholas T. Williams, Paweł Morzywołek, and Kara E. Rudolph. Modified treatment policies that depend on the natural history of treatment, 2026.

Jianqing Fan and Irene Gijbels. *Local Polynomial Modelling and Its Applications*. Chapman and Hall, London, 1996. ISBN 9780412983214.

Stéphane Gaïffas. Convergence rates for pointwise curve estimation with a degenerate design. *Mathematical Methods of Statistics*, 14:1–27, 2005.

Stéphane Gaïffas. On pointwise adaptive curve estimation based on inhomogeneous data. *ESAIM: Probability and Statistics*, 11:344–364, 2007. doi: 10.1051/ps:2007023.

Sebastian Haneuse and Andrea Rotnitzky. Estimation of the effect of interventions that modify the received treatment. *Statistics in Medicine*, 32(30):5260–5277, 2013. doi: 10.1002/sim.5907.

Keisuke Hirano and Guido W. Imbens. The propensity score with continuous treatments. In Andrew Gelman and Xiao-Li Meng, editors, *Applied Bayesian Modeling and Causal Inference from Incomplete-Data Perspectives*, pages 73–84. Wiley, Chichester, 2005. doi: 10.1002/0470090456.ch7.

Wassily Hoeffding. Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association*, 58(301):13–30, 1963. doi: 10.1080/01621459.1963.10500830.

Katherine L. Hoffman, Diego Salazar-Barreto, Nicholas T. Williams, Kara E. Rudolph, and Iván Díaz. Studying continuous, time-varying, and/or complex exposures using longitudinal modified treatment policies. *Epidemiology*, 35(5):667–675, 2024. doi: 10.1097/EDE.0000000000001764.

Kosuke Imai and David A van Dyk. Causal inference with general treatment regimes. *Journal of the American Statistical Association*, 99(467):854–866, 2004. doi: 10.1198/016214504000001187.

Guido W. Imbens. The role of the propensity score in estimating dose-response functions. *Biometrika*, 87(3):706–710, 2000. doi: 10.1093/biomet/87.3.706.

Olav Kallenberg. *Foundations of Modern Probability*. Springer, New York, 2 edition, 2002. doi: 10.1007/978-1-4757-4015-8.

Edward H. Kennedy, Zongming Ma, Matthew D. McHugh, and Dylan S. Small. Nonparametric methods for doubly robust estimation of continuous treatment effects. *Journal of the Royal Statistical Society: Series B*, 79(4):1229–1245, 2017. doi: 10.1111/rssb.12212.

Taehyeon Koo, Elizabeth A. Stuart, Kara E. Rudolph, and Caleb H. Miles. Causal effects of modified treatment policies under positivity violations: A partial identification approach, 2026.

Mark G. Low. On nonparametric confidence intervals. *The Annals of Statistics*, 25(6):2547–2554, 1997. doi: 10.1214/aos/1030741084.

David McCoy, Wenxin Zhang, Alan Hubbard, Mark J. van der Laan, and Alejandro Schuler. Data-adaptive identification of effect modifiers through stochastic shift interventions and cross-validated targeted learning. *Biostatistics*, 27:kxag020, 2026. doi: 10.1093/biostatistics/kxag020.

Iván Díaz Muñoz and Mark van der Laan. Population intervention causal effects based on stochastic interventions. *Biometrics*, 68(2):541–549, 2011. doi: 10.1111/j.1541-0420.2011.01685.x.

Judea Pearl. *Causality: Models, Reasoning, and Inference*. Cambridge University Press, Cambridge, 2 edition, 2009. ISBN 9780521895606.

James M. Robins, Miguel A. Hernán, and Babette Brumback. Marginal structural models and causal inference in epidemiology. *Epidemiology*, 11(5):550–560, 2000. doi: 10.1097/00001648-200009000-00011.

Charles J. Stone. Optimal global rates of convergence for nonparametric regression. *The Annals of Statistics*, 10(4):1040–1053, 1982. doi: 10.1214/aos/1176345969.

Alexandre B. Tsybakov. *Introduction to Nonparametric Estimation*. Springer, New York, 2009. doi: 10.1007/b13794.

Lars van der Laan, Wenbo Zhang, and Peter B. Gilbert. Nonparametric estimation of the causal effect of a stochastic threshold-based intervention. *Biometrics*, 79(3):1868–1881, 2022. doi: 10.1111/biom.13690.

Mark J. van der Laan and James M. Robins. *Unified Methods for Censored Longitudinal Data and Causality*. Springer, New York, 2003. doi: 10.1007/978-0-387-21700-0.

Mark J. van der Laan and Sherri Rose. *Targeted Learning*. Springer, New York, 2011. doi: 10.1007/978-1-4419-9782-1.

Nicholas T. Williams and Iván Díaz. lmtp: An R package for estimating the causal effects of modified treatment policies. *Observational Studies*, 9(2):103–122, 2023. doi: 10.1353/obs.2023.0019.

Jessica G. Young, Miguel A. Hernán, and James M. Robins. Identification, estimation and approximation of risk under interventions that depend on the natural value of treatment using observational data. *Epidemiologic Methods*, 3(1):1–19, 2014. doi: 10.1515/em-2012-0001.