Proceedings of the National Academy of Sciences

Supplementary Information for

Accuracy Limits of Causal Trees for Individualized Treatment Effects

Matias D. Cattaneo, Jason M. Klusowski, and Ruiqi (Rae) Yu

Corresponding authors are Matias D. Cattaneo (cattaneo@princeton.edu), Jason M. Klusowski (jason.klusowski@princeton.edu), and Ruiqi (Rae) Yu (rae.yu@princeton.edu).

November 16, 2026

Contents of this supplement

Supporting information text
Supporting figures and SI references

Supporting Information Text

Contents

SA-1 Overview and Omitted Details 4
- SA-1.1 Notation . . . 4
- SA-1.2 Proof Strategy of Theorem 1 . . . 5
- SA-1.3 Probability Bounds for Integrated Error . . . 6
- SA-1.4 Regularization and Small Cells . . . 6
- SA-1.5 α-Regularity and Small Subgroups . . . 7

SA-2 Regression Estimation 7
- SA-2.1 No Sample Splitting . . . 8
- SA-2.2 Sample Splitting . . . 10
- SA-2.3 Unbiasedness . . . 10

SA-3 Causal Effect Estimation 11
- SA-3.1 IPW Estimator . . . 12
  - SA-3.1.1 No Sample Splitting . . . 13
  - SA-3.1.2 Sample Splitting . . . 13
- SA-3.2 DIM Estimator . . . 14
  - SA-3.2.1 No Sample Splitting . . . 14
  - SA-3.2.2 Sample Splitting . . . 16
- SA-3.3 SSE Estimator . . . 16
  - SA-3.3.1 No Sample Splitting . . . 16
  - SA-3.3.2 Sample Splitting . . . 19
- SA-3.4 Squared T-Statistic Estimators . . . 19
- SA-3.5 Unbiasedness . . . 20

SA-4 Correction to Eicker (1979) 21
- SA-4.1 Proof of Lemma SA-35 . . . 21

SA-5 Proofs 22
- SA-5.1 Technical Lemmas . . . 22
- SA-5.2 Proof of Theorem SA-1 . . . 31
- SA-5.3 Proof of Theorem SA-2 . . . 33
- SA-5.4 Proof of Theorem SA-3 . . . 35
- SA-5.5 Proof of Theorem SA-4 . . . 35
- SA-5.6 Proof of Theorem SA-5 . . . 36
- SA-5.7 Proof of Theorem SA-6 . . . 38
- SA-5.8 Proof of Theorem SA-7 . . . 39
- SA-5.9 Proof of Lemma SA-8 . . . 56
- SA-5.10 Proof of Corollary SA-9 . . . 56
- SA-5.11 Proof of Corollary SA-10 . . . 56
- SA-5.12 Proof of Corollary SA-11 . . . 56
- SA-5.13 Proof of Corollary SA-12 . . . 56
- SA-5.14 Proof of Corollary SA-13 . . . 56
- SA-5.15 Proof of Corollary SA-14 . . . 57
- SA-5.16 Proof of Corollary SA-15 . . . 57
- SA-5.17 Proof of Lemma SA-16 . . . 57
- SA-5.18 Proof of Lemma SA-17 . . . 58
- SA-5.19 Proof of Theorem SA-18 . . . 59
- SA-5.20 Proof of Theorem SA-19 . . . 60
- SA-5.21 Proof of Theorem SA-20 . . . 61
- SA-5.22 Proof of Theorem SA-21 . . . 64
- SA-5.23 Proof of Theorem SA-22 . . . 65
- SA-5.24 Proof of Theorem SA-23 . . . 67
- SA-5.25 Proof of Lemma SA-24 . . . 68
- SA-5.26 Proof of Lemma SA-25 . . . 68
- SA-5.27 Proof of Lemma SA-26 . . . 69
- SA-5.28 Proof of Theorem SA-27 . . . 71

SA-5.29 Proof of Lemma SA-28 . . . 72
SA-5.30 Proof of Theorem SA-29 . . . 74
SA-5.31 Proof of Theorem SA-30 . . . 74
SA-5.32 Proof of Theorem SA-31 . . . 75
SA-5.33 Proof of Theorem SA-32 . . . 75
SA-5.34 Proof of Theorem SA-33 . . . 76
SA-5.35 Proof of Lemma SA-34 . . . 76

SA-6 Numerical Evidence **77**
SA-6.1 Numerical Implementation . . . 78
SA-6.2 Simulations . . . 78
SA-6.3 Empirical Resampling . . . 78

References **79**

SA-1. Overview and Omitted Details

This supplement gives the proofs of the main results, additional theoretical results, and numerical evidence. Section SA-2 first shows that CART regression trees fail to attain any polynomial rate of uniform convergence even when the conditional mean is constant. Section SA-3 then treats causal effect estimation, with inverse probability weighting (IPW), difference in means (DIM), and sum of squared errors (SSE) estimators in Sections SA-3.1, SA-3.2, and SA-3.3, respectively. Here CART denotes classification and regression trees, CATE denotes conditional average treatment effect, NSS denotes no sample splitting, and HON denotes honesty. The results needed for the main paper are listed below.

- **Proof of Theorem 1.** Apply Corollary SA-11, Corollary SA-13, Theorem SA-20, Theorem SA-22, Theorem SA-30, and Theorem SA-32. For deeper NSS-DIM and NSS-SSE trees, the theorems below give weaker depth conditions and an explicit relation between depth and the probability lower bound. The depth condition in the main paper implies their $a = b/2$ specialization. The depth condition is, however, unnecessary for IPW, honest trees, and stumps. For NSS-IPW trees and NSS-DIM stumps, we obtain the limiting lower probability bound b. SSE has the same exact limit for split locations, but its error bound also includes the constant needed to pass from the bivariate split score to the terminal CATE contrast.

- **Proof of Theorem 2.** Apply Corollary SA-12, Corollary SA-14, Theorem SA-21, Theorem SA-23, Theorem SA-31, and Theorem SA-33.

- **Proof of Remark 1.** Apply Corollary SA-15 for uniform inconsistency and Corollary SA-14 for integrated consistency.

Theorem 1 of the main paper uses C_1, C_2 for no sample splitting and C_3, C_4 for honesty. Below, constants are defined separately for each result, often as C_1, C_2; the honest uniform lower bounds give explicit admissible values of C_3, C_4. All four criteria satisfy the exact root split limits in Theorems SA-1, SA-18, and SA-27, and Corollary SA-9. The supplemental probability bounds, in turn, imply those stated in the main paper. These limits, however, concern the root split, so they do not give exact uniform-error probabilities after arbitrary further splitting. For deeper NSS-DIM and NSS-SSE trees, Lemma SA-49 bounds the largest terminal error using a remainder of order $J \log n/M$ for a root child of size $M \geq C_0 \log n$ with J terminal descendants. This gives the sharper supplemental depth conditions while retaining Theorem 1 of the main paper as a consequence.

We also give the following results and details.

- Correction to Eicker (1979). We correct the normalization in [1, Theorem 5] and give the corresponding vector Ornstein–Uhlenbeck (O-U) limit in Section SA-4.

- Probability bounds corresponding to the integrated error bounds in Theorem 2, in Section SA-1.3.

- Uniform inconsistency of maximally grown honest univariate regression trees at depth of order $\log \log n$ under Assumption SA-1, in Theorem SA-7, and its HON-IPW consequence in Corollary SA-15. Among the uniform-error lower bounds, only these depth-specific results require maximal growth.

- The proof strategy for Theorem 1 and a discussion of regularization, small cells, and α-regularity, in Sections SA-1.2, SA-1.4, and SA-1.5.

- Simulations and an empirical resampling study, with designs and implementation details, in Section SA-6.

- Unbiasedness results for regression and causal trees in Sections SA-2.3 and SA-3.5, respectively, and squared t-statistic estimators in Section SA-3.4.

SA-1.1. Notation.

We use the following notation.

- **Sets.** $\mathbb{R}$ is the set of real numbers and $\mathbb{N}$ the positive integers. For $n \in \mathbb{N}$ we write $[n] = \{1, \ldots, n\}$.

- **Vectors and matrices.** Boldface lowercase letters (e.g. $\mathbf{x}$) denote column vectors, and boldface uppercase letters (e.g. $\mathbf{A}$) denote matrices or, when specified, random vectors and vector-valued processes. For a vector $\mathbf{x}$, its i-th component is x_i. I_d is the $d \times d$ identity matrix. Write $\mathbf{X} = (x_{i,j})_{i \in [n], j \in [p]}$ for the covariate matrix and, in the causal model, $\mathbf{d} = (d_1, \ldots, d_n)^\top$ for the assignment vector.

- **Norms.** For $\mathbf{x} \in \mathbb{R}^d$, define $\|\mathbf{x}\| = (\sum_{i=1}^d x_i^2)^{1/2}$, and $\|\mathbf{x}\|_\infty = \max_{1 \leq i \leq d} |x_i|$. For a matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$, the operator norm is $\|\mathbf{A}\| = \sup_{\mathbf{x} \in \mathbb{R}^n,\ \|\mathbf{x}\|=1} \|\mathbf{A}\mathbf{x}\|$. For a bounded measurable function $g : \mathcal{A} \to \mathbb{R}$, $\|g\|_\infty = \sup_{x \in \mathcal{A}} |g(x)|$. For the covariate vector $\mathbf{x}_i$, let P_X denote its marginal distribution and let F_X denote its distribution function. For $1 \leq q < \infty$, the population L^q norm of a scalar function g is $\|g\|_q = (\int_{\mathcal{X}} |g(\mathbf{x})|^q \, \mathrm{d}F_X(\mathbf{x}))^{1/q}$.

- **Asymptotics.** We use $\to$ for ordinary convergence and $\to_{\mathbb{P}}$ for convergence in probability. For real sequences, $a_n = o(b_n)$ means $\limsup_{n \to \infty} |a_n|/|b_n| = 0$, while $a_n = O(b_n)$ means there exist constants C and $N > 0$ such that $|a_n| \leq C|b_n|$ for all $n > N$. For sequences of random variables, $a_n = o_{\mathbb{P}}(b_n)$ means $|a_n|/|b_n| \to_{\mathbb{P}} 0$, while $a_n = O_{\mathbb{P}}(b_n)$ means $\limsup_{M \to \infty} \limsup_{n \to \infty} \mathbb{P}(|a_n/b_n| \geq M) = 0$. We also write $a_n \asymp b_n$ for two-sided bounds on positive sequences, $a_n \sim b_n$ when $a_n/b_n \to 1$, and $\rightsquigarrow$ for convergence in distribution. $a_n \uparrow a$ and $a_n \downarrow a$ denote monotone convergence. Unless indicated

otherwise, log is the natural logarithm. An $o(1)$ in a deterministic threshold denotes a deterministic sequence tending to zero. A lower bound with threshold $\sqrt{(2+o(1))\log\log n}$ holds for some such sequence, possibly depending on the fixed parameters in the statement. We do not require the bound to hold for every vanishing sequence. Random remainders, by contrast, are denoted by $o_{\mathbb{P}}(1)$ or $O_{\mathbb{P}}(1)$. Constants are local to each statement, and C in a proof denotes a generic positive constant that may change from line to line. The covariate dimension p is fixed throughout. Unless stated otherwise, generic positive constants may depend on p and on the distributional quantities listed in the statement. We give sharp constants explicitly, as in the split probabilities below.

- **Probability.** We write $\mathbb{P}$, $\mathbb{E}$, and $\mathbb{V}$ for probability, expectation, and variance. Subscripts on $\mathbb{P}$, $\mathbb{E}$, and $\mathbb{V}$ indicate the underlying law or the starting point of a Markov process. For a random element Z, $\mathcal{L}(Z)$ denotes its probability law, and $\mathcal{L}(Z \mid \mathscr{G})$ denotes its conditional law given a σ-field $\mathscr{G}$. The notation $Z \sim P_Z$ means $\mathcal{L}(Z) = P_Z$, and $Z \overset{d}{=} W$ means $\mathcal{L}(Z) = \mathcal{L}(W)$. Products of probability laws are denoted by $\otimes$ or $\bigotimes$. Throughout, auxiliary variables and couplings may be constructed on extensions of the probability space that preserve the joint law of the original variables. All trees are binary. All rules for constructing trees and selecting partitions are assumed to be measurable. Under honesty, observation indices are local to each fold (subsample).

- **Other.** $\mathbb{1}(\cdot)$ denotes the indicator function. For real numbers a, b, $a \wedge b = \min\{a, b\}$, $a \vee b = \max\{a, b\}$, and $a_+ = \max\{a, 0\}$. For σ-fields, $\mathscr{F} \vee \mathscr{G} = \sigma(\mathscr{F} \cup \mathscr{G})$. The relation $A \succ 0$ means that the symmetric matrix A is positive definite, and $\Longrightarrow$ denotes logical implication. For two random variables X and Y, $X \perp\!\!\!\perp Y$ means X and Y are independent. For $x \in \mathbb{R}$, $\lfloor x \rfloor$ and $\lceil x \rceil$ denote the floor and ceiling of x, respectively. $\mathsf{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ denotes the Gaussian distribution with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. $\mathsf{Beta}(\alpha, \beta)$ denotes the Beta distribution with parameters (α, β). A stochastic process $\{B(t), 0 \le t \le 1\}$ is a Brownian bridge if B is a continuous Gaussian process with $\mathbb{E}[B(t)] = 0$ and $\mathbb{E}[B(t)B(s)] = \min\{t, s\} - ts$. When real sequences such as r_n or s_n are used as split index cutoffs, inequalities such as $r_n \le k \le n - r_n$ mean $\lceil r_n \rceil \le k \le n - \lceil r_n \rceil$. This rounding convention is suppressed below. When a split index optimization is written as an optimization over $k \in [n]$, the endpoint $k = n$ is not a valid split and the optimization is understood over the valid split indices $1 \le k < n$, together with any validity restrictions specific to the estimator. A maximum of nonnegative scores or approximation errors over an empty candidate set is defined as zero, unless a local comparison of winning candidates explicitly uses the value $-\infty$.

- **Boundary sequences.** In the normalized coordinates described after Assumption SA-1, for $a \in (0, 1)$ and a deterministic sequence $\eta_n \downarrow 0$, define

$$\mathcal{X}_n(a, \eta_n) = \left\{ \mathbf{x} \in [0,1]^p : x_j \le \eta_n n^{a-1} \text{ or } 1 - x_j \le \eta_n n^{a-1} \text{ for some } j \in [p] \right\}.$$

SA-1.2. Proof Strategy of Theorem 1. The causal model and estimators are defined in Assumption SA-2 and Definitions SA-4–SA-6. Our lower bounds follow from root splits that leave few observations on one side. Let $M = n$ without sample splitting and $M = n_{\mathsf{T}}$ for honest construction. For each $l \in \{\texttt{IPW}, \texttt{DIM}, \texttt{SSE}\}$, write $\hat{\imath}$ for the number of construction observations in the left root child and $\hat{\jmath}$ for the split coordinate. For $0 < a < b < 1$ and $j \in [p]$, we have $\mathbb{P}(\hat{\jmath} = j) \to 1/p$ and

$$\lim_{n\to\infty} \mathbb{P}\big(M^a \le \hat{\imath} \le M^b \mid \hat{\jmath} = j\big) = \frac{b-a}{2} \qquad \text{and} \qquad \lim_{n\to\infty} \mathbb{P}\big(M - M^b \le \hat{\imath} \le M - M^a \mid \hat{\jmath} = j\big) = \frac{b-a}{2}. \tag{SA-1}$$

Consequently, for each $j \in [p]$,

$$\frac{\log \min\{\hat{\imath}, M - \hat{\imath}\}}{\log M} \,\Big|\, \hat{\jmath} = j \rightsquigarrow \mathsf{Uniform}(0,1) \qquad \text{and} \qquad \liminf_{n\to\infty} \mathbb{P}\big(\min\{\hat{\imath}, M - \hat{\imath}\} \le M^b,\ \hat{\jmath} = j\big) \ge \frac{b}{p}.$$

Under the fold-size conditions for honest lower bounds, M is comparable to n. Averaging over the small child then gives the stump lower bound.

For NSS regression and IPW, a root-child average is a convex combination of its terminal-descendant averages. The largest absolute terminal error is therefore at least the absolute root-child error. DIM and SSE, however, require an additional bound on the descendant treatment-arm weights, giving the depth conditions in Theorems SA-20 and SA-30. In particular,

$$K = o(\log n) \quad \Longrightarrow \quad 2^K \log n = o\Big(n^{b/4}\sqrt{\log\log n}\Big) \qquad \text{for every fixed } b > 0.$$

The uniform lower bounds thus rule out every polynomial rate. Regression, IPW, and honest trees, by contrast, need no such depth restriction. For honest trees, occupancy and anti-concentration on the independent estimation fold replace the comparison of construction-sample averages.

The root-split argument has four steps, from criterion approximation to the limiting split probabilities.

Step 1. Split criterion approximation. The IPW criterion is exactly CART applied to $y_i(d_i - \xi)/\{\xi(1-\xi)\}$. The DIM proxy instead uses

$$\frac{d_i \varepsilon_i(1)}{\xi} - \frac{(1-d_i)\varepsilon_i(0)}{1-\xi}.$$

The SSE proxy, in turn, is the sum of the CART criteria for $d_i\varepsilon_i(1)/\xi$ and $(1-d_i)\varepsilon_i(0)/(1-\xi)$, weighted by ξ and $1-\xi$. The arm means cancel from within-arm contrasts. A dyadic product bound controls replacement of empirical arm counts. Partial-sum bounds also exclude cuts with fewer than $r_n = \lceil (\log n)^8 \rceil$ observations on either side.

Step 2. One-coordinate strong approximation. The scalar approximation of [2] couples ordered residual sums to Brownian motion. For SSE, conditional on the covariate ordering and treatment assignments, the observed residuals have the product law of two independent treatment-arm streams. This remains true when the two potential errors of an individual are dependent, since only one is observed. Coupling each stream and subtracting its endpoint interpolation then gives two independent standard Brownian bridges.

Step 3. Fixed-dimensional covariates. Retain the first and last $\lfloor n/(\log n)^4 \rfloor$ observations in each coordinate ordering. Replace repeated occurrences across these arrays by independent copies. A conditional martingale bound then gives an $o_{\mathbb{P}}(1)$ change in squared scores. The resulting boundary arrays are independent, and the one-coordinate result excludes the omitted central ranges. This gives the joint boundary limit for fixed p.

Step 4. Joint block maxima. A logarithmic time change turns a standardized Brownian bridge into a stationary Ornstein–Uhlenbeck (O-U) process. Lemma SA-35 then rescales the O-U norm limit of [3]. For unequal arm variances, Lemma SA-41 combines the fixed-interval tail of [4] with the Dirichlet characterization in [5]. Lemma SA-36, in turn, uses the Markov property and separating gaps of length $o(\log n)$ to obtain independent limiting block maxima. Their common scale and length-proportional intensities give

$$\mathbb{P}(\text{a given block contains the limiting maximum}) = \frac{\text{block length}}{\text{total length}}.$$

Dividing the joint winning-block probabilities by $\mathbb{P}(\hat{\jmath} = j) \to 1/p$ yields the conditional limits in (SA-1). Continuity of the limiting laws also makes the empirical tie rule immaterial.

SA-1.3. Probability Bounds for Integrated Error. Theorem 2 in the main text bounds moments of the population L^q norm, for every fixed $1 \le q < \infty$. The corresponding probability bounds use the same assumptions and deterministic depth bound $K = K_n$. For no sample splitting, Corollary SA-12, Theorem SA-21, and Theorem SA-31 imply that, for $n \ge 2$ and $l \in \{\texttt{IPW}, \texttt{DIM}, \texttt{SSE}\}$,

$$\mathbb{P}\left(\|\hat{\tau}_l - \tau\|_q \ge C_1 \left(\frac{2^K \log^{\max\{3,q+1\}}(n)}{n} \right)^{1/\max\{2,q\}} \right) \le C_3 n^{-6},$$

where C_1 and C_3 depend only on q, the fixed dimension p, and the distributional quantities in the corresponding result. For honest sample splitting, Corollary SA-14, Theorem SA-23, and Theorem SA-33 imply that, for every $\delta \in (0,1)$,

$$\mathbb{P}\left(\|\check{\tau}_l - \tau\|_q \ge C_2 \left(\frac{2^K}{(n_\tau + 1)\delta} \right)^{1/\max\{2,q\}} \right) \le \delta,$$

where C_2 depends on q and the outcome/treatment distribution, but not on p or the fold sizes. The tree is constructed using only the construction fold, and no restriction on the relative fold sizes is required. Constants, in either case, need not be uniform in q.

Without sample splitting, a sufficient condition for L^q consistency is $2^{K_n} \log^3 n/n \to 0$ when $1 \le q \le 2$, or $2^{K_n} \log^{q+1} n/n \to 0$ when $q > 2$. For honest sample splitting, $2^{K_n}/(n_\tau + 1) \to 0$ suffices for every fixed finite $q \ge 1$.

SA-1.4. Regularization and Small Cells. Minimum node-size constraints and related penalties can reduce the variance from small cells, but they can also change the chosen splits. Minimum node sizes are imposed in the `causalTree` and `htetree` implementations [6, 7]. Small cells can, however, be useful when the conditional mean or treatment effect changes sharply. They can also arise from sampling noise when the target is locally flat. The data alone may not distinguish these cases.

For example, the `causalTree` documentation defines `minsize` as requiring at least `minsize` treated and `minsize` control observations in each leaf during tree construction, with default `minsize`=2 and example code setting `minsize`=20 [6, `causalTree`, Arguments and Examples]. The `htetree` package, in turn, uses default `minsize`=20 in its high-level causal tree, IPW, and matching wrappers. Its `hte_forest` wrapper uses the same default for a causal tree that defines subgroups for summarizing an already fitted forest; this setting does not control the constituent forest trees [7, version 0.1.23, pp. 19–24 and `R/hte_forest.R`]. These defaults restrict tree construction but do not guarantee the same minimum treatment-arm counts in an independent estimation sample.

A minimum node-size rule can also prevent splits that would recover the treatment effect after an earlier poor split. Consider, for illustration, the threshold CATE function

$$\tau(\mathbf{x}) = \mathbb{1}(x_1 \le 1/2).$$

Suppose the tree first splits on an irrelevant coordinate at $x_2 = 0.05$, leaving most observations in $\{x_2 > 0.05\}$. The threshold in x_1 remains in that descendant, so a later split can recover it there. Recovery over the full support also requires an appropriate fit in the other descendant. A minimum node-size requirement can, however, force the first split to be more balanced, leaving fewer observations in the larger descendant than after the original imbalanced split. Thus an imbalanced split can harm pointwise estimation and still help a deeper tree recover the treatment effect. The corresponding recovery argument for regression trees is discussed in [8, Section 1].

SA-1.5. α-Regularity and Small Subgroups. α-regularity requires every split to leave at least an α fraction of the parent observations in each child node. This sample-balance requirement is part of [9, Definitions 4 and 4b], [10, Assumption 3], and [11, Specification 1]. A related condition balances the population volumes of the children [12, Assumption A2]. The counts refer to the sample on which balance is imposed; for the double-sample trees in [9], this is the estimation sample. Sample balance excludes highly imbalanced splits in that sample. Population balance, in turn, can require more splits to isolate small subgroups, as the following example shows.

Consider

$$\tau(\mathbf{x}) = \mathbb{1}(x_1 \leq \eta), \qquad \eta \in (0,1),$$

with $\mathbf{x} = (x_1, \ldots, x_p)^\top$ uniformly distributed on $[0,1]^p$. At the population level, an unrestricted axis-aligned tree can represent this function exactly with the single split $x_1 = \eta$. In this illustration, α-regularity refers to the conditional population mass of each child given its parent. If $\eta < \alpha$, an α-regular tree cannot make that split at the root. To isolate the subgroup exactly, any terminal cell contained in $\{x_1 \leq \eta\}$ must have at least

$$\left\lceil \log_{1/\alpha}(1/\eta) \right\rceil$$

splits along x_1 in its ancestry. After k such splits, its width in that coordinate is at least α^k, so containment requires $\alpha^k \leq \eta$. This bound, however, concerns exact isolation. It does not specify the depth needed for approximate recovery. Requiring more splits also makes the tree less parsimonious and less adaptive in its choice of variables. When p is large relative to n, each additional split creates another opportunity to select an irrelevant covariate. A balanced tree may, nevertheless, approximate the treatment effect in integrated loss without recovering the subgroup itself.

SA-2. Regression Estimation

We give a self-contained extension of the regression results in the 2022 version of [13]. Besides their interest for regression estimation, these results provide the arguments used for causal trees in Section SA-3.

The earlier work develops the scalar split argument and discusses higher dimensions in Section 4.4 [13]. Under Assumption SA-1, we establish the joint limiting law of the selected split coordinate and location in any fixed dimension $p \geq 1$, and use it to obtain lower bounds for stumps and deeper trees. For honest trees, our probability bounds use one construction fold and one independent estimation fold. The earlier work instead bounded the expected error of honest stumps and used a fresh independent sample at every level to study deeper honest trees [13, Sections 4.5 and 5]. We also bound the integrated L^q error for every fixed $1 \leq q < \infty$ without sample splitting. Section SA-3 applies these arguments to IPW, DIM, and SSE causal trees, with and without sample splitting.

For univariate CART decision stumps, [14, Theorem 3.1] obtained a cube-root convergence rate under smoothness and identification conditions. [15, Theorem 2.1 and the discussion following it] gave the joint limit for the split and fitted leaf means, correcting the earlier limiting distribution while retaining the cube-root rate. Our lower bounds, by contrast, concern constant regression functions, which belong to standard smoothness classes but do not identify an interior population split. Sampling fluctuations then determine the empirical split, which can lie near the boundary with probability bounded away from zero. Our results are therefore consistent with the classical cube-root analyses and show why their identification conditions matter. Pointwise results for regression functions that identify an interior split do not give uniform guarantees over classes containing flat functions.

Consider the canonical regression model where the observed data $\{(y_i, \mathbf{x}_i^\top) : i = 1, 2, \ldots, n\}$ is a random sample satisfying

$$y_i = \mu(\mathbf{x}_i) + \varepsilon_i, \qquad \mathbb{E}[\varepsilon_i \mid \mathbf{x}_i] = 0, \tag{SA-2}$$

with $\mathbf{x}_i = (x_{i,1}, x_{i,2}, \ldots, x_{i,p})^\top$ a vector of p covariates taking values on some support set $\mathcal{X}$.

Assumption SA-1 (Location Regression Model)**.** *$\mathcal{D} = \{(y_i, \mathbf{x}_i^\top) : 1 \leq i \leq n\}$ is generated by i.i.d. pairs $(\mathbf{x}_i, \varepsilon_i)$ satisfying Equation (SA-2) and the following conditions for all $i = 1, 2, \cdots, n$.*

1. *$y_i = \mu(\mathbf{x}_i) + \varepsilon_i$, with $\mathbb{E}[\varepsilon_i \mid \mathbf{x}_i] = 0$ and $\mathbf{x}_i \perp\!\!\!\perp \varepsilon_i$.*
2. *$\mu(\mathbf{x}) = \mu^\star$ for all $\mathbf{x} \in \mathcal{X} \subseteq \mathbb{R}^p$, where $\mu^\star$ is some constant.*
3. *$x_{i,1}, \ldots, x_{i,p}$ are independent and continuously distributed.*
4. *There exists $\chi > 0$ such that $\mathbb{E}[\exp(\lambda \varepsilon_i)] < \infty$ for all $|\lambda| < 1/\chi$ and $\sigma^2 = \mathbb{E}[\varepsilon_i^2] > 0$.*

Let F_j be the continuous marginal distribution function of $x_{i,j}$. The probability integral transform [16, Chapter II, Theorem 2.1, p. 28] makes the coordinates $F_j(x_{i,j})$ independent uniforms and preserves the observed coordinate ranks in both construction and estimation samples almost surely. Also, almost surely, $x_j \leq x_{i,j}$ if and only if $F_j(x_j) \leq F_j(x_{i,j})$, simultaneously for every real x_j and every observed threshold $x_{i,j}$. Indeed, the sampled uniform CDF values avoid the at most countably many levels at which F_j is constant on a nontrivial interval. With observed thresholds and the coordinate/cut-rank tie rule below, the transformation therefore preserves the candidate child memberships, selected split ranks, and terminal estimates.

We evaluate predictions on $[0,1]^p$, the closed support of the transformed law, and retain the notation $\mathbf{x}$ for its points. At the boundary, including $\mathbf{0}$ and points without an original-scale preimage, predictions use the same $\le$ and $>$ split inequalities. Every terminal leaf of the prescribed trees contains interior points of the cube, so evaluating on the closed cube preserves the supremum error for the constant target. The transformation also preserves integrated risk. Statements over $\mathcal{X}_n(a,\eta_n)$ apply to every point of this normalized set, whose boundary bands are interpreted on the original scale through $F_j(x_j)$. The covariate dimension p is fixed throughout the asymptotic analysis, and constants in the upper bounds without sample splitting may depend on it.

Given any tree T, the CART estimator is defined as follows.

Definition SA-1 (CART Estimate)**.** *Given a tree T and an estimation sample $\mathcal{D}_\mu = \{(y_i, \mathbf{x}_i^\top) : i = 1,2,\dots,n_\mu\}$, with $n_\mu \le n$, let $\mathtt{t}$ be the unique terminal node in T containing $\mathbf{x} \in \mathcal{X}$. Using only $\mathcal{D}_\mu$ for the sums and counts, define the CART estimator by*

$$\hat{\mu}(\mathbf{x}; \mathsf{T}, \mathcal{D}_\mu) = \frac{1}{n(\mathtt{t})} \sum_{i:\mathbf{x}_i \in \mathtt{t}} y_i,$$

where $n(\mathtt{t}) = \sum_{i=1}^{n_\mu} \mathbb{1}(\mathbf{x}_i \in \mathtt{t})$ is the "local" sample size. In case $n(\mathtt{t}) = 0$, take $\hat{\mu}(\mathbf{x}; \mathsf{T}, \mathcal{D}_\mu) = 0$.

Definition SA-2 (Tree Construction)**.** *Given a dataset $\mathcal{D}_\mathsf{T} = \{(y_i, \mathbf{x}_i^\top) : i = 1,2,\dots,n_\mathsf{T}\}$, with $n_\mathsf{T} \le n$, a parent node $\mathtt{t}$ in the tree (i.e., a region in $\mathcal{X}$) is divided into two child nodes, $\mathtt{t}_\mathtt{L}$ and $\mathtt{t}_\mathtt{R}$, by minimizing the sum of squared errors (SSE),*

$$\min_{1\le j\le p} \min_{\beta_\mathtt{L},\beta_\mathtt{R},\varsigma\in\mathbb{R}} \sum_{\mathbf{x}_i \in \mathtt{t}} \left(y_i - \beta_\mathtt{L}\mathbb{1}(x_{i,j} \le \varsigma) - \beta_\mathtt{R}\mathbb{1}(x_{i,j} > \varsigma)\right)^2, \qquad \text{[SA-3]}$$

where $(\beta_\mathtt{L}, \beta_\mathtt{R}, \varsigma, j)$ denote the two child-node outputs, split point, and split coordinate, respectively. We optimize over splits with at least one construction-sample observation in each child and leave a parent terminal if no such split exists. Each candidate threshold is the largest observed construction-sample value of the split coordinate in its left child. The cut rank is the number of construction observations in the left child. Among tied optimal splits, choose the smallest coordinate index and then the smallest cut rank within that coordinate. Denote the resulting CART tree by $\mathsf{T}(\mathcal{D}_\mathsf{T})$. Throughout, the root has depth zero, each child is one level deeper than its parent, and tree depth is the maximum node depth.

Definition SA-3 (Sample Splitting)**.** *Apply Definitions SA-1 and SA-2 to the available random sample $\mathcal{D} = \{(y_i, \mathbf{x}_i^\top) : i = 1,2,\dots,n\}$ as follows.*

- *No Sample Splitting (NSS). The dataset $\mathcal{D}$ is used for both the tree construction and the regression mean estimation, that is, $\mathcal{D}_\mathsf{T} = \mathcal{D}$ and $\mathcal{D}_\mu = \mathcal{D}$. The CART tree estimator is then*

$$\hat{\mu}^{\mathtt{NSS}}(\mathbf{x}) = \hat{\mu}(\mathbf{x}; \mathsf{T}(\mathcal{D}), \mathcal{D}).$$

- *Honesty (HON). The dataset $\mathcal{D}$ is divided into two independent datasets $\mathcal{D}_\mathsf{T}$ and $\mathcal{D}_\mu$ with positive sample sizes n_T and n_μ satisfying $n_\mathsf{T} + n_\mu = n$. The CART tree estimator is*

$$\hat{\mu}^{\mathtt{HON}}(\mathbf{x}) = \hat{\mu}(\mathbf{x}; \mathsf{T}(\mathcal{D}_\mathsf{T}), \mathcal{D}_\mu).$$

SA-2.1. No Sample Splitting.

We first study the root split without sample splitting (NSS).

Decision Stumps. For each variable $j = 1,2,\dots,p$, let π_j be the permutation such that $x_{\pi_j(i),j}$ is nondecreasing in the index $i = 1,2,\dots,n$. Minimizing Equation (SA-3) is equivalent to maximizing the *impurity gain*

$$\sum_{\mathbf{x}_i\in\mathtt{t}} \left(y_i - \overline{y}_\mathtt{t}\right)^2 - \sum_{\mathbf{x}_i\in\mathtt{t}} \left(y_i - \overline{y}_{\mathtt{t}_\mathtt{L}}\mathbb{1}(\mathbf{x}_i \in \mathtt{t}_\mathtt{L}) - \overline{y}_{\mathtt{t}_\mathtt{R}}\mathbb{1}(\mathbf{x}_i \in \mathtt{t}_\mathtt{R})\right)^2 = \frac{\left(\frac{1}{\sqrt{n(\mathtt{t})}}\sum_{\mathbf{x}_i\in\mathtt{t}_\mathtt{L}}(y_i-\mu) - \frac{n(\mathtt{t}_\mathtt{L})}{n(\mathtt{t})}\frac{1}{\sqrt{n(\mathtt{t})}}\sum_{\mathbf{x}_i\in\mathtt{t}}(y_i-\mu)\right)^2}{(n(\mathtt{t}_\mathtt{L})/n(\mathtt{t}))(1-n(\mathtt{t}_\mathtt{L})/n(\mathtt{t}))},$$

where $\overline{y}_\mathtt{t} = n(\mathtt{t})^{-1}\sum_{\mathbf{x}_i\in\mathtt{t}} y_i$. The impurity gain can equivalently be written as

$$\frac{n(\mathtt{t}_\mathtt{L})n(\mathtt{t}_\mathtt{R})}{n(\mathtt{t})}\left(\overline{y}_{\mathtt{t}_\mathtt{L}} - \overline{y}_{\mathtt{t}_\mathtt{R}}\right)^2.$$

For a tree of depth one ($K = 1$), or decision stump, either objective selects a splitting coordinate $\hat{\jmath}$ and a splitting index $\hat{\imath}$, with child nodes

$$\mathtt{t}_\mathtt{L} = \{\mathbf{u} \in \mathcal{X} : u_{\hat{\jmath}} \le x_{\pi_{\hat{\jmath}}(\hat{\imath}),\hat{\jmath}}\} \qquad \text{and} \qquad \mathtt{t}_\mathtt{R} = \{\mathbf{u} \in \mathcal{X} : u_{\hat{\jmath}} > x_{\pi_{\hat{\jmath}}(\hat{\imath}),\hat{\jmath}}\}.$$

The tree output can then be written as

$$\hat{\mu}^{\mathtt{stump}}(\mathbf{x}) = \begin{cases} \overline{y}_{\mathtt{t}_\mathtt{L}}, & \mathbf{x} \in \mathtt{t}_\mathtt{L}, \\ \overline{y}_{\mathtt{t}_\mathtt{R}}, & \mathbf{x} \in \mathtt{t}_\mathtt{R} \end{cases}.$$

The following theorem now gives the probabilities of root splits near the boundary of $\mathcal{X}$. It also describes the child sample sizes, since the selected rank $\hat{\imath}$ is the number of observations in the left child. The ordered split point is $x_{\pi_{\hat{\jmath}}(\hat{\imath}),\hat{\jmath}}$, so $\hat{\imath} = \#\{\mathbf{x}_i : x_{i,\hat{\jmath}} \le x_{\pi_{\hat{\jmath}}(\hat{\imath}),\hat{\jmath}}\}$.

Theorem SA-1 (Imbalanced Splits)**.** *Suppose Assumption SA-1 holds, and let $(\hat{\imath}, \hat{\jmath})$ be the CART split index and split coordinate at the root node. Any measurable rule may resolve ties. For each $a, b \in (0,1)$ with $a < b$, and $\ell \in [p]$, we have*

$$\lim_{n\to\infty} \mathbb{P}\big(n^a \le \hat{\imath} \le n^b,\ \hat{\jmath} = \ell\big) = \frac{b-a}{2p} \qquad \text{and} \qquad \lim_{n\to\infty} \mathbb{P}\big(n - n^b \le \hat{\imath} \le n - n^a,\ \hat{\jmath} = \ell\big) = \frac{b-a}{2p}.$$

For each $\ell \in [p]$, $\mathbb{P}(\hat{\jmath} = \ell) \to 1/p$, and hence

$$\lim_{n\to\infty} \mathbb{P}\big(n^a \le \hat{\imath} \le n^b \mid \hat{\jmath} = \ell\big) = \frac{b-a}{2} \qquad \text{and} \qquad \lim_{n\to\infty} \mathbb{P}\big(n - n^b \le \hat{\imath} \le n - n^a \mid \hat{\jmath} = \ell\big) = \frac{b-a}{2}.$$

Consequently,

$$\frac{\log \min\{\hat{\imath}, n - \hat{\imath}\}}{\log n} \,\Big|\, \hat{\jmath} = \ell \ \rightsquigarrow\ \mathsf{Uniform}(0,1).$$

The same window and uniform limits hold without conditioning.

We next translate this split behavior into bounds on estimation error.

Theorem SA-2 (Convergence Rates for Decision Stumps)**.** *Suppose Assumption SA-1 holds and the CART tree has depth $K = 1$. Then $\sup_{\mathbf{x}\in\mathcal{X}} |\hat{\mu}^{\texttt{stump}}(\mathbf{x}) - \mu(\mathbf{x})| \to_{\mathbb{P}} 0$, and for any $a, b \in (0,1)$ with $a < b$, we have*

$$\lim_{n\to\infty} \mathbb{P}\left(\sup_{\mathbf{x}\in\mathcal{X}} |\hat{\mu}^{\texttt{stump}}(\mathbf{x}) - \mu(\mathbf{x})| \ge \sigma n^{-b/2} \sqrt{2 \log\log(n)} \right) = b, \tag{SA-4}$$

and, under the probability integral transform normalization described after Assumption SA-1, for any deterministic sequence $\eta_n \downarrow 0$,

$$\liminf_{n\to\infty} \inf_{\mathbf{x}\in\mathcal{X}_n(a,\eta_n)} \mathbb{P}\Big(|\hat{\mu}^{\texttt{stump}}(\mathbf{x}) - \mu(\mathbf{x})| \ge \sigma n^{-b/2} \sqrt{(2 + o(1)) \log\log(n)} \Big) \ge \frac{b-a}{2p}. \tag{SA-5}$$

Deep Trees. The stump lower bound also extends to any finite recursive tree with at least one split, even when its depth varies with n.

Theorem SA-3 (Convergence Rates for Deep Trees)**.** *Suppose Assumption SA-1 holds, and the CART tree has depth at least one. Then for any $b \in (0,1)$, we have*

$$\liminf_{n\to\infty} \mathbb{P}\left(\sup_{\mathbf{x}\in\mathcal{X}} |\hat{\mu}^{\texttt{NSS}}(\mathbf{x}) - \mu(\mathbf{x})| \ge \sigma n^{-b/2} \sqrt{(2 + o(1)) \log\log(n)} \right) \ge b.$$

Thus, CART trees need not attain any polynomial rate in n when convergence is required uniformly over the support $\mathcal{X}$ and over the admissible data generating processes.

For every fixed finite $q \ge 1$, integrated L^q error can still be small because the cells responsible for the uniform-error lower bounds have small P_X measure.

Theorem SA-4 (L^q Convergence Rate for NSS)**.** *Suppose Assumption SA-1 holds, and fix $1 \le q < \infty$. For $n \ge 2$ and any data-dependent axis-aligned tree with nonempty sample leaves and depth at most a deterministic $K = K_n$, the NSS regression estimator satisfies*

$$\big(\mathbb{E}[\|\hat{\mu}^{\texttt{NSS}} - \mu\|_q^q]\big)^{1/q} \le C_1 \left(\frac{2^K \log^{\max\{3,q+1\}}(n)}{n} \right)^{1/\max\{2,q\}}.$$

Moreover,

$$\mathbb{P}\left(\|\hat{\mu}^{\texttt{NSS}} - \mu\|_q \ge C_2 \left(\frac{2^K \log^{\max\{3,q+1\}}(n)}{n} \right)^{1/\max\{2,q\}} \right) \le C_3 n^{-6}.$$

The positive constants C_1, C_2, and C_3 depend only on q, the fixed dimension p, and the distribution of ε_i.

SA-2.2. Sample Splitting. Under sample splitting, we again bound uniform error from below and integrated L^q error from above for every fixed $1 \leq q < \infty$.

Theorem SA-5 (Uniform Rates for HON)**.** *Suppose Assumption SA-1 holds, and the construction fold honest CART tree has at least one split. Assume $\rho \leq n_{\mathcal{T}}/n_\mu \leq \rho^{-1}$ for some $\rho \in (0,1)$. Then for any $b \in (0,1)$, we have*

$$\liminf_{n\to\infty} \mathbb{P}\left(\sup_{\mathbf{x}\in\mathcal{X}} |\hat{\mu}^{\texttt{HON}}(\mathbf{x}) - \mu(\mathbf{x})| \geq \frac{\rho\sigma}{2\sqrt{1+\rho}} n^{-b/2} \right) \geq \frac{\rho\sigma^4 b}{16(1+\rho)\max\{3\sigma^4, \mathbb{E}[\varepsilon_i^4]\}}.$$

Theorem SA-6 (L^q Convergence Rate for HON)**.** *Suppose Assumption SA-1 holds, and fix $1 \leq q < \infty$. For any regression tree constructed using only $\mathcal{D}_{\mathcal{T}}$ and having depth at most a deterministic $K = K_n$, the honest estimator satisfies*

$$\left(\mathbb{E}[\|\hat{\mu}^{\texttt{HON}} - \mu\|_q^q]\right)^{1/q} \leq C_1 \left(\frac{2^K}{n_\mu + 1}\right)^{1/\max\{2,q\}},$$

where $C_1, C_2 > 0$ depend only on q and the distribution of y_i, not on p or the fold sizes. Moreover, for every $\delta \in (0,1)$,

$$\mathbb{P}\left(\|\hat{\mu}^{\texttt{HON}} - \mu\|_q \geq C_2 \left(\frac{2^K}{(n_\mu+1)\delta}\right)^{1/\max\{2,q\}} \right) \leq \delta.$$

Compared with Theorem SA-3, the uniform-error lower bound in Theorem SA-5 lacks the factor $\sqrt{(2+o(1))\log\log(n)}$. The moment bound in Theorem SA-6, in turn, contains no logarithmic factors and depends on the estimation-fold size. To prove it, we condition on the construction partition and sum the expected risk over its at most 2^K leaves. For no sample splitting, these bounds imply L^q consistency if $2^K \log^3 n/n \to 0$ for $1 \leq q \leq 2$, or $2^K \log^{q+1} n/n \to 0$ for $q > 2$. Honest L^q consistency, by contrast, follows whenever $2^K/(n_\mu+1) \to 0$, for every fixed finite $q \geq 1$.

For a scalar covariate, maximally grown honest regression trees are uniformly inconsistent at depth of order $\log\log n$ under Assumption SA-1. With the root assigned depth zero, a maximal CART tree splits every node containing at least two construction observations until the deterministic depth cap $K_{n_{\mathcal{T}}}$. The splits follow the criterion, observed-value thresholds, and tie rule in Definition SA-2, including when the gain is zero. No additional minimum child size or pruning is imposed.

Theorem SA-7 (Uniform Inconsistency for Univariate HON Regression Trees)**.** *Suppose Assumption SA-1 holds with $p = 1$, and $\rho \leq n_{\mathcal{T}}/n_\mu \leq \rho^{-1}$ for a fixed $\rho \in (0,1)$. Grow a maximal CART tree on the construction fold and use the independent estimation fold for its terminal averages. There are constants $C_1, C_2, \epsilon > 0$, depending only on the fixed distribution of y_i and on ρ, such that, for every deterministic integer sequence $(K_{n_{\mathcal{T}}})$ with $K_{n_{\mathcal{T}}} \geq C_1 \log\log n_{\mathcal{T}}$ eventually,*

$$\liminf_{n_{\mathcal{T}}\to\infty} \mathbb{P}\left(\sup_{\mathbf{x}\in\mathcal{X}} \left|\hat{\mu}^{\texttt{HON}}(\mathbf{x}) - \mu(\mathbf{x})\right| > \epsilon \right) \geq C_2.$$

SA-2.3. Unbiasedness. For honest regression CART, only empty estimation cells contribute to the bias, through the zero estimate assigned to them. Without sample splitting, by contrast, the same residuals determine the partition and the terminal averages. We therefore assume that the full residual vector is symmetric and that reversing its sign leaves the complete selected partition unchanged.

Lemma SA-8 (Regression unbiasedness)**.** *Suppose Assumption SA-1 holds, and fix $\mathbf{x} \in \mathcal{X}$. Let $N_{\texttt{HON}}(\mathbf{x})$ denote the final estimation sample size in the terminal node containing $\mathbf{x}$ for the honest CART estimator. Then*

$$\mathbb{E}[\hat{\mu}^{\texttt{HON}}(\mathbf{x})] = \mu(\mathbf{x}) - \mu(\mathbf{x})\,\mathbb{P}(N_{\texttt{HON}}(\mathbf{x}) = 0).$$

For the estimator without sample splitting, if in addition the residual distribution is symmetric about zero, $\varepsilon_i \overset{d}{=} -\varepsilon_i$, and the final partition is unchanged by the simultaneous reflection $y_i \mapsto 2\mu - y_i$ for all i, with covariates fixed, then

$$\mathbb{E}[\hat{\mu}^{\texttt{NSS}}(\mathbf{x})] = \mu(\mathbf{x}).$$

Here, partition invariance includes every tie-breaking, stopping, refinement, and pruning decision. The prescribed coordinate/cut-rank rule, for example, preserves the selected split when the relevant residuals change sign. The whole partition is also preserved if every eligible node is split up to a predetermined maximum depth and eligibility depends only on covariates, node counts, or invariant squared split scores. Determinism alone is, however, insufficient, since a rule may depend on the signs of the responses. These invariance and symmetry assumptions are used only for regression NSS unbiasedness, whereas the honest formulas use independent estimation.

SA-3. Causal Effect Estimation

We now turn to heterogeneous causal effect estimation, using the assumptions and causal tree definitions from the main paper, repeated below.

Assumption SA-2 (Data Generating Process). *$\mathcal{D} = \{(y_i, d_i, \mathbf{x}_i^\top) : 1 \le i \le n\}$ is generated by i.i.d. vectors $(\mathbf{x}_i, d_i, \varepsilon_i(0), \varepsilon_i(1))$, where $y_i = d_i y_i(1) + (1-d_i)y_i(0)$, $\mathbf{x}_i = (x_{i,1}, \ldots, x_{i,p})^\top$, and the following conditions hold for all $d = 0, 1$ and $i = 1, 2, \ldots, n$.*

(i) *$d_i \perp\!\!\!\perp (\mathbf{x}_i, \varepsilon_i(0), \varepsilon_i(1))$, and $\xi = \mathbb{P}(d_i = 1) \in (0, 1)$.*

(ii) *$y_i(d) = \mu_d(\mathbf{x}_i) + \varepsilon_i(d)$, with $\mathbb{E}[\varepsilon_i(d)] = 0$ and $\mathbf{x}_i \perp\!\!\!\perp (\varepsilon_i(0), \varepsilon_i(1))$.*

(iii) *$\mu_d(\mathbf{x}) = \mu_d^\star$ for all $\mathbf{x} \in \mathcal{X}$, where $\mu_d^\star$ is some constant, and $\mathcal{X}$ is the support of $\mathbf{x}_i$.*

(iv) *$x_{i,1}, \ldots, x_{i,p}$ are independent and continuously distributed.*

(v) *There exists $\chi > 0$ such that $\mathbb{E}[\exp(\lambda \varepsilon_i(d))] < \infty$ for all $|\lambda| < 1/\chi$ and $\mathbb{E}[\varepsilon_i^2(d)] > 0$.*

For $d \in \{0, 1\}$, write $\xi_d = \mathbb{P}(d_i = d)$, so that $\xi_1 = \xi$ and $\xi_0 = 1 - \xi$. Throughout this section, $\tau(\mathbf{x}) \equiv \tau^\star = \mu_1^\star - \mu_0^\star$ denotes the constant treatment effect.

Statements involving $\mathcal{X} = [0, 1]^p$ or $\mathcal{X}_n(a, \eta_n)$ use the marginal probability integral transforms and closed evaluation domain described after Assumption SA-1. For the prescribed causal trees, these transforms preserve the ranks and the routing at observed thresholds almost surely. They also leave unchanged the joint law of treatment assignments and potential outcome errors, their independence from the covariates, the constant arm means, and the CATE.

The causal tree estimators are constructed as follows.

Definition SA-4 (CATE Estimators). *Given a tree T and an estimation sample $\mathcal{D}_\tau = \{(y_i, d_i, \mathbf{x}_i^\top) : i = 1, 2, \ldots, n_\tau\}$, with $n_\tau \le n$, let t be the unique terminal node in T containing $\mathbf{x} \in \mathcal{X}$.*

- *The Inverse Probability Weighting (IPW) estimator is then*

$$\hat{\tau}_{\mathtt{IPW}}(\mathbf{x}; \mathsf{T}, \mathcal{D}_\tau) = \frac{1}{n(\mathsf{t})} \sum_{i:\mathbf{x}_i \in \mathsf{t}} \frac{d_i - \xi}{\xi(1-\xi)} y_i,$$

  where $n(\mathsf{t}) = \sum_{i=1}^{n_\tau} \mathbb{1}(\mathbf{x}_i \in \mathsf{t})$ is the "local" sample size. In case $n(\mathsf{t}) = 0$, take $\hat{\tau}_{\mathtt{IPW}}(\mathbf{x}; \mathsf{T}, \mathcal{D}_\tau) = 0$.

- *The Difference in Means (DIM) estimator is*

$$\hat{\tau}_{\mathtt{DIM}}(\mathbf{x}; \mathsf{T}, \mathcal{D}_\tau) = \frac{1}{n_1(\mathsf{t})} \sum_{i:\mathbf{x}_i \in \mathsf{t}} d_i y_i - \frac{1}{n_0(\mathsf{t})} \sum_{i:\mathbf{x}_i \in \mathsf{t}} (1 - d_i) y_i,$$

  where $n_d(\mathsf{t}) = \sum_{i=1}^{n_\tau} \mathbb{1}(\mathbf{x}_i \in \mathsf{t}, d_i = d)$, for $d = 0, 1$, are the "local" sample sizes. In case $n_0(\mathsf{t}) = 0$ or $n_1(\mathsf{t}) = 0$, take $\hat{\tau}_{\mathtt{DIM}}(\mathbf{x}; \mathsf{T}, \mathcal{D}_\tau) = 0$.

Because these estimators are constant on each terminal node t, write

$$\hat{\tau}_l(\mathsf{t}; \mathsf{T}, \mathcal{D}_\tau) = \hat{\tau}_l(\mathbf{x}; \mathsf{T}, \mathcal{D}_\tau), \qquad \mathbf{x} \in \mathsf{t}, \quad l \in \{\mathtt{IPW}, \mathtt{DIM}\}.$$

The same convention applies to the abbreviated NSS and HON estimators defined below.

Definition SA-5 (Tree Construction). *Use the sample $\mathcal{D}_\mathsf{T} = \{(y_i, d_i, \mathbf{x}_i^\top) : i = 1, 2, \ldots, n_\mathsf{T}\}$, with $n_\mathsf{T} \le n$, to construct the tree T. All sums, counts, and fitted child means below accordingly use $\mathcal{D}_\mathsf{T}$. We optimize over candidate splits with at least one construction-sample observation in each child. For criteria involving treatment-arm means or node-specific treatment coefficients, both treatment-arm denominators must also be positive in each child. A parent with no valid split remains terminal. Each candidate threshold is the largest observed construction-sample value of the split coordinate in its left child. Among tied optimal splits, choose the smallest coordinate index and then the smallest cut rank within that coordinate.*

- *Variance Maximization. A parent node t (i.e., a terminal node of the current tree) in a previous tree T' is divided into two child nodes, $\mathsf{t}_\mathtt{L}$ and $\mathsf{t}_\mathtt{R}$, forming the new tree T, by maximizing*

$$\frac{n(\mathsf{t}_\mathtt{L}) n(\mathsf{t}_\mathtt{R})}{n(\mathsf{t})} \Big(\hat{\tau}_l(\mathsf{t}_\mathtt{L}; \mathsf{T}, \mathcal{D}_\mathsf{T}) - \hat{\tau}_l(\mathsf{t}_\mathtt{R}; \mathsf{T}, \mathcal{D}_\mathsf{T}) \Big)^2, \qquad l \in \{\mathtt{IPW}, \mathtt{DIM}\}. \tag{SA-6}$$

  The resulting causal trees are denoted by $\mathsf{T}_{\mathtt{IPW}}(\mathcal{D}_\mathsf{T})$ and $\mathsf{T}_{\mathtt{DIM}}(\mathcal{D}_\mathsf{T})$, respectively, for $l \in \{\mathtt{IPW}, \mathtt{DIM}\}$.

- *SSE Minimization. A parent node t (i.e., a terminal node of the current tree) in the previous tree T' is divided into two child nodes, $\mathsf{t}_\mathtt{L}$ and $\mathsf{t}_\mathtt{R}$, forming the next tree T, by solving*

$$\min_{a_\mathtt{L}, b_\mathtt{L}, a_\mathtt{R}, b_\mathtt{R} \in \mathbb{R}} \sum_{\mathbf{x}_i \in \mathsf{t}_\mathtt{L}} (y_i - a_\mathtt{L} - b_\mathtt{L} d_i)^2 + \sum_{\mathbf{x}_i \in \mathsf{t}_\mathtt{R}} (y_i - a_\mathtt{R} - b_\mathtt{R} d_i)^2, \tag{SA-7}$$

  where only the data $\mathcal{D}_\mathsf{T}$ is used. The valid coordinate and threshold minimize this profiled objective. The resulting causal tree is denoted by $\mathsf{T}_{\mathtt{SSE}}(\mathcal{D}_\mathsf{T})$.

Definition SA-6 (Sample Splitting and Estimators)**.** *Apply Definitions SA-4 and SA-5 to the available random sample* $\mathcal{D} = \{(y_i, d_i, \mathbf{x}_i^\top) : i = 1, 2, \ldots, n\}$ *as follows.*

- *No Sample Splitting (NSS). The dataset* $\mathcal{D}$ *is used for both the tree construction and the treatment effect estimation, that is,* $\mathcal{D}_{\mathsf{T}} = \mathcal{D}$ *and* $\mathcal{D}_\tau = \mathcal{D}$*. The causal tree estimators are then*

$$\begin{aligned}\hat{\tau}_{\texttt{IPW}}(\mathbf{x}) &= \hat{\tau}_{\texttt{IPW}}(\mathbf{x}; \mathsf{T}_{\texttt{IPW}}(\mathcal{D}), \mathcal{D}),\\ \hat{\tau}_{\texttt{DIM}}(\mathbf{x}) &= \hat{\tau}_{\texttt{DIM}}(\mathbf{x}; \mathsf{T}_{\texttt{DIM}}(\mathcal{D}), \mathcal{D}), \quad \textit{and}\\ \hat{\tau}_{\texttt{SSE}}(\mathbf{x}) &= \hat{\tau}_{\texttt{DIM}}(\mathbf{x}; \mathsf{T}_{\texttt{SSE}}(\mathcal{D}), \mathcal{D}),\end{aligned}$$

- *Honesty (HON). The dataset* $\mathcal{D}$ *is divided into two independent datasets* $\mathcal{D}_{\mathsf{T}}$ *and* $\mathcal{D}_\tau$ *with positive sample sizes* n_{T} *and* n_τ *satisfying* $n_{\mathsf{T}} + n_\tau = n$*. The causal tree estimators are*

$$\begin{aligned}\check{\tau}_{\texttt{IPW}}(\mathbf{x}) &= \hat{\tau}_{\texttt{IPW}}(\mathbf{x}; \mathsf{T}_{\texttt{IPW}}(\mathcal{D}_{\mathsf{T}}), \mathcal{D}_\tau),\\ \check{\tau}_{\texttt{DIM}}(\mathbf{x}) &= \hat{\tau}_{\texttt{DIM}}(\mathbf{x}; \mathsf{T}_{\texttt{DIM}}(\mathcal{D}_{\mathsf{T}}), \mathcal{D}_\tau), \quad \textit{and}\\ \check{\tau}_{\texttt{SSE}}(\mathbf{x}) &= \hat{\tau}_{\texttt{DIM}}(\mathbf{x}; \mathsf{T}_{\texttt{SSE}}(\mathcal{D}_{\mathsf{T}}), \mathcal{D}_\tau).\end{aligned}$$

The estimators $\hat{\tau}_l(\mathbf{x})$ and $\check{\tau}_l(\mathbf{x})$, $l \in \{\texttt{IPW}, \texttt{DIM}, \texttt{SSE}\}$, depend on tree depth. We suppress this dependence except where a result requires an explicit depth restriction, and retain K in the integrated-risk bounds.

Root-split convention. Throughout the lower bounds, a requirement of depth one or at least one split means that the root is split according to the prescribed criterion whenever a valid split exists. Otherwise the root remains terminal. For construction-sample size $M = n_{\mathsf{T}} \geq 2$, let π_ℓ order the sample along coordinate ℓ, and define

$$\mathcal{V}_{\texttt{DIM}}(M) = \mathcal{V}_{\texttt{SSE}}(M) = \left\{(k, \ell) \in [M-1] \times [p] : \sum_{i=1}^{k} \mathbb{1}(d_{\pi_\ell(i)} = d) > 0,\ \sum_{i=k+1}^{M} \mathbb{1}(d_{\pi_\ell(i)} = d) > 0 \quad (d = 0, 1)\right\}.$$

For regression and IPW, every element of $[M-1] \times [p]$ is valid almost surely. For DIM and SSE, the cut $k = \lfloor M/2 \rfloor$ in coordinate one is valid whenever each child contains both arms. The ordered treatment indicators are independent $\mathsf{Bernoulli}(\xi)$ variables, so

$$\mathbb{P}\big(\mathcal{V}_{\texttt{DIM}}(M) = \varnothing\big) \leq \xi^k + (1-\xi)^k + \xi^{M-k} + (1-\xi)^{M-k} \leq 2\big\{\xi^{\lfloor M/2 \rfloor} + (1-\xi)^{\lfloor M/2 \rfloor}\big\} \to 0.$$

On $\{\mathcal{V}_{\texttt{DIM}}(M) = \varnothing\}$, assign the auxiliary rank and coordinate $(1, 1)$ and leave the root terminal. The corresponding threshold defines auxiliary candidate regions and logarithmic split statistics. Actual root children are used only on $\{\mathcal{V}_{\texttt{DIM}}(M) \neq \varnothing\}$. For every event A,

$$0 \leq \mathbb{P}(A) - \mathbb{P}\big(A \cap \{\mathcal{V}_{\texttt{DIM}}(M) \neq \varnothing\}\big) \leq \mathbb{P}\big(\mathcal{V}_{\texttt{DIM}}(M) = \varnothing\big) = o(1).$$

Thus the convention leaves all asymptotic lower bounds unchanged. The midpoint cut, however, proves existence only. The selected cut still optimizes the prescribed criterion. Write $\mathcal{V}_l = \mathcal{V}_l(n)$ for $l \in \{\texttt{DIM}, \texttt{SSE}\}$.

In the DIM and SSE arguments, call a sample rectangle or tree node $\mathtt{t}$ valid if $n_0(\mathtt{t}) \wedge n_1(\mathtt{t}) > 0$, with counts taken in the sample under consideration. For such $\mathtt{t}$, write

$$\bar{\varepsilon}_{1,\mathtt{t}} = \frac{1}{n_1(\mathtt{t})} \sum_{\mathbf{x}_i \in \mathtt{t}} d_i \varepsilon_i(1), \qquad \bar{\varepsilon}_{0,\mathtt{t}} = \frac{1}{n_0(\mathtt{t})} \sum_{\mathbf{x}_i \in \mathtt{t}} (1-d_i) \varepsilon_i(0) \qquad \text{and} \qquad \Delta(\mathtt{t}) = \bar{\varepsilon}_{1,\mathtt{t}} - \bar{\varepsilon}_{0,\mathtt{t}}.$$

SA-3.1. IPW Estimator. Under Assumption SA-2, the transformed outcomes are i.i.d. and

$$\mathbb{E}\left[y_i \frac{d_i - \xi}{\xi(1-\xi)} \,\middle|\, \mathbf{x}_i\right] = \mathbb{E}[y_i(1) - y_i(0) \mid \mathbf{x}_i] = \mu_1 - \mu_0 = \tau.$$

Their centered residuals satisfy

$$\tilde{\varepsilon}_i = y_i \frac{d_i - \xi}{\xi(1-\xi)} - \tau = \frac{d_i \varepsilon_i(1)}{\xi} - \frac{(1-d_i)\varepsilon_i(0)}{1-\xi} + (d_i - \xi)\left(\frac{\mu_1}{\xi} + \frac{\mu_0}{1-\xi}\right). \tag{SA-8}$$

Then $\tilde{\varepsilon}_i \perp\!\!\!\perp \mathbf{x}_i$ and $\mathbb{E}[\tilde{\varepsilon}_i] = 0$. Conditioning on d_i gives

$$\mathbb{E}[e^{\lambda \tilde{\varepsilon}_i}] = \xi e^{\lambda(\mu_1/\xi - \tau)} \mathbb{E}[e^{(\lambda/\xi)\varepsilon_i(1)}] + (1-\xi) e^{\lambda(-\mu_0/(1-\xi) - \tau)} \mathbb{E}[e^{-\lambda \varepsilon_i(0)/(1-\xi)}]. \tag{SA-9}$$

For $\beta = 2\chi / \min\{\xi, 1-\xi\}$, each expectation on the right is finite whenever $|\lambda| \leq 1/\beta$, by Assumption SA-2. The conditional-variance formula also yields

$$\mathbb{V}[\tilde{\varepsilon}_i] = \frac{\mathbb{E}[\varepsilon_i(1)^2]}{\xi} + \frac{\mathbb{E}[\varepsilon_i(0)^2]}{1-\xi} + \frac{\{(1-\xi)\mu_1 + \xi\mu_0\}^2}{\xi(1-\xi)} > 0. \tag{SA-10}$$

The transformed regression model therefore satisfies Assumption SA-1. Since its CART criterion and terminal average are exactly the IPW criterion and estimator, the results in Section SA-2 give the following corollaries.

SA-3.1.1. No Sample Splitting. Let $(\hat{\imath}, \hat{\jmath})$ denote the cut rank and coordinate selected by the IPW root split.

Corollary SA-9 (Imbalanced Split)**.** *Suppose Assumption SA-2 holds. Then for each $a, b \in (0,1)$ with $a < b$, for every $\ell \in [p]$,*

$$\lim_{n\to\infty} \mathbb{P}\big(n^a \le \hat{\imath} \le n^b,\ \hat{\jmath} = \ell\big) = \frac{b-a}{2p} \qquad \text{and} \qquad \lim_{n\to\infty} \mathbb{P}\big(n - n^b \le \hat{\imath} \le n - n^a,\ \hat{\jmath} = \ell\big) = \frac{b-a}{2p}.$$

For each $\ell \in [p]$, $\mathbb{P}(\hat{\jmath} = \ell) \to 1/p$. Conditional on $\hat{\jmath} = \ell$, both window probabilities converge to $(b-a)/2$.

Corollary SA-10 (Uniform Rates for Decision Stumps)**.** *Suppose Assumption SA-2 holds, and the tree has depth $K = 1$. Then $\sup_{\mathbf{x}\in\mathcal{X}} |\hat{\tau}_{\texttt{IPW}}(\mathbf{x}) - \tau(\mathbf{x})| \to_{\mathbb{P}} 0$, and for any $a, b \in (0,1)$ with $a < b$, we have*

$$\lim_{n\to\infty} \mathbb{P}\left(\sup_{\mathbf{x}\in\mathcal{X}} |\hat{\tau}_{\texttt{IPW}}(\mathbf{x}) - \tau(\mathbf{x})| \ge \sigma n^{-b/2} \sqrt{2 \log\log(n)} \right) = b,$$

where $\sigma^2 = \mathbb{V}\left[\frac{d_i y_i(1)}{\xi} - \frac{(1-d_i) y_i(0)}{1-\xi}\right]$. Moreover, under the marginal probability integral transform normalization described after Assumption SA-2, for any deterministic sequence $\eta_n \downarrow 0$,

$$\liminf_{n\to\infty} \inf_{\mathbf{x}\in\mathcal{X}_n(a,\eta_n)} \mathbb{P}\Big(|\hat{\tau}_{\texttt{IPW}}(\mathbf{x}) - \tau(\mathbf{x})| \ge \sigma n^{-b/2} \sqrt{(2+o(1)) \log\log(n)} \Big) \ge \frac{b-a}{2p}.$$

Corollary SA-11 (Uniform Rates for Deep Trees)**.** *Suppose Assumption SA-2 holds. Then for any $b \in (0,1)$ and any NSS-IPW transformed outcome tree with at least one split, we have*

$$\liminf_{n\to\infty} \mathbb{P}\left(\sup_{\mathbf{x}\in\mathcal{X}} |\hat{\tau}_{\texttt{IPW}}(\mathbf{x}) - \tau(\mathbf{x})| \ge \sigma n^{-b/2} \sqrt{\log\log(n)} \right) \ge b,$$

where $\sigma^2 = \mathbb{V}[y_i(d_i - \xi)/\{\xi(1-\xi)\}]$.

Corollary SA-12 (L^q Convergence Rate for NSS)**.** *Suppose Assumption SA-2 holds, and fix $1 \le q < \infty$. For $n \ge 2$ and any data-dependent axis-aligned tree with nonempty sample leaves and depth at most a deterministic $K = K_n$, the NSS-IPW terminal estimator satisfies*

$$\left(\mathbb{E}[\|\hat{\tau}_{\texttt{IPW}} - \tau\|_q^q]\right)^{1/q} \le C_1 \left(\frac{2^K \log^{\max\{3,q+1\}}(n)}{n} \right)^{1/\max\{2,q\}}.$$

Moreover,

$$\mathbb{P}\left(\|\hat{\tau}_{\texttt{IPW}} - \tau\|_q \ge C_2 \left(\frac{2^K \log^{\max\{3,q+1\}}(n)}{n} \right)^{1/\max\{2,q\}} \right) \le C_3 n^{-6}.$$

The positive constants C_1, C_2, and C_3 depend only on q, the fixed dimension p, and the distribution of $\tilde{\varepsilon}_i = y_i(d_i - \xi)/\{\xi(1-\xi)\} - \tau$.

SA-3.1.2. Sample Splitting.

Corollary SA-13 (Uniform Rates for HON)**.** *Suppose Assumption SA-2 holds, and the construction fold HON-IPW tree has at least one split. Assume $\rho \le n_{\mathsf{T}}/n_\tau \le \rho^{-1}$ for some $\rho \in (0,1)$. Then for any $b \in (0,1)$, we have*

$$\liminf_{n\to\infty} \mathbb{P}\left(\sup_{\mathbf{x}\in\mathcal{X}} |\check{\tau}_{\texttt{IPW}}(\mathbf{x}) - \tau(\mathbf{x})| \ge \frac{\rho\sigma}{2\sqrt{1+\rho}} n^{-b/2} \right) \ge \frac{\rho\sigma^4 b}{16(1+\rho)\max\{3\sigma^4, \mathbb{E}[\tilde{\varepsilon}_i^4]\}},$$

where $\sigma^2 = \mathbb{V}[\tilde{\varepsilon}_i]$ and $\tilde{\varepsilon}_i = y_i \frac{d_i - \xi}{\xi(1-\xi)} - \tau$.

Corollary SA-14 (L^q Convergence Rate for HON)**.** *Suppose Assumption SA-2 holds, and fix $1 \le q < \infty$. For any axis-aligned tree constructed using only $\mathcal{D}_{\mathsf{T}}$ and having depth at most a deterministic $K = K_n$, the honest IPW terminal estimator satisfies*

$$\left(\mathbb{E}[\|\check{\tau}_{\texttt{IPW}} - \tau\|_q^q]\right)^{1/q} \le C_1 \left(\frac{2^K}{n_\tau + 1} \right)^{1/\max\{2,q\}},$$

where $C_1, C_2 > 0$ depend only on q and τ and the distribution of $\tilde{\varepsilon}_i = y_i(d_i - \xi)/\{\xi(1-\xi)\} - \tau$, not on p or the fold sizes. Moreover, for every $\delta \in (0,1)$,

$$\mathbb{P}\left(\|\check{\tau}_{\texttt{IPW}} - \tau\|_q \ge C_2 \left(\frac{2^K}{(n_\tau+1)\delta} \right)^{1/\max\{2,q\}} \right) \le \delta.$$

Corollary SA-15 (Uniform Inconsistency for Univariate HON-IPW Trees). *Suppose Assumption SA-2 holds with $p = 1$. Assume $\rho \le n_{\mathsf{T}}/n_\tau \le \rho^{-1}$ for a fixed $\rho \in (0,1)$. Grow a maximal IPW tree on the construction fold, following the growth convention of Section SA-2.2 with the IPW rule in Definition SA-5, and use the independent estimation fold for its terminal estimates. There are constants $C_1, C_2, \epsilon > 0$, depending only on the fixed distribution of $(y_i(0), y_i(1), d_i)$ and on ρ, such that, for every deterministic integer sequence $(K_{n_{\mathsf{T}}})$ with $K_{n_{\mathsf{T}}} \ge C_1 \log\log n_{\mathsf{T}}$ eventually,*

$$\liminf_{n_{\mathsf{T}}\to\infty} \mathbb{P}\left(\sup_{\mathbf{x}\in\mathcal{X}} \left| \check{\tau}_{\mathtt{IPW}}(\mathbf{x}) - \tau(\mathbf{x}) \right| > \epsilon \right) \ge C_2.$$

SA-3.2. DIM Estimator. The `DIM` estimator does not admit the same exact CART representation. We instead approximate its split criterion, uniformly over indices and coordinates, by the CART criterion for $d_i\varepsilon_i(1)/\xi - (1-d_i)\varepsilon_i(0)/(1-\xi)$.

SA-3.2.1. No Sample Splitting.

Approximation Results on Decision Stumps. Let π_ℓ be the permutation of $[n]$ such that $x_{\pi_\ell(1),\ell} \le x_{\pi_\ell(2),\ell} \le \cdots \le x_{\pi_\ell(n),\ell}$, $1 \le \ell \le p$. For a root split with k observations in the left child along coordinate ℓ, write the DIM criterion and its IPW proxy based on potential-outcome residuals as

$$\mathscr{I}^{\mathtt{DIM}}(k,\ell) = \frac{k(n-k)}{n}\left(\hat{\tau}_{\mathtt{L}}^{\mathtt{DIM}}(k,\ell) - \hat{\tau}_{\mathtt{R}}^{\mathtt{DIM}}(k,\ell)\right)^2 \qquad \text{and} \qquad \bar{\mathscr{I}}^{\mathtt{IPW}}(k,\ell) = \frac{k(n-k)}{n}\left(\bar{\tau}_{\mathtt{L}}^{\mathtt{IPW}}(k,\ell) - \bar{\tau}_{\mathtt{R}}^{\mathtt{IPW}}(k,\ell)\right)^2,$$

where

$$\begin{aligned}
\hat{\tau}_{\mathtt{L}}^{\mathtt{DIM}}(k,\ell) &= \frac{\sum_{i=1}^{k} d_{\pi_\ell(i)} y_{\pi_\ell(i)}}{\sum_{i=1}^{k} d_{\pi_\ell(i)}} - \frac{\sum_{i=1}^{k} (1-d_{\pi_\ell(i)}) y_{\pi_\ell(i)}}{\sum_{i=1}^{k} (1-d_{\pi_\ell(i)})}, \\
\hat{\tau}_{\mathtt{R}}^{\mathtt{DIM}}(k,\ell) &= \frac{\sum_{i=k+1}^{n} d_{\pi_\ell(i)} y_{\pi_\ell(i)}}{\sum_{i=k+1}^{n} d_{\pi_\ell(i)}} - \frac{\sum_{i=k+1}^{n} (1-d_{\pi_\ell(i)}) y_{\pi_\ell(i)}}{\sum_{i=k+1}^{n} (1-d_{\pi_\ell(i)})}, \\
\bar{\tau}_{\mathtt{L}}^{\mathtt{IPW}}(k,\ell) &= \frac{1}{k}\sum_{i=1}^{k} \frac{d_{\pi_\ell(i)}}{\xi}\varepsilon_{\pi_\ell(i)}(1) - \frac{1}{k}\sum_{i=1}^{k} \frac{1-d_{\pi_\ell(i)}}{1-\xi}\varepsilon_{\pi_\ell(i)}(0), \\
\bar{\tau}_{\mathtt{R}}^{\mathtt{IPW}}(k,\ell) &= \frac{1}{n-k}\sum_{i=k+1}^{n} \frac{d_{\pi_\ell(i)}}{\xi}\varepsilon_{\pi_\ell(i)}(1) - \frac{1}{n-k}\sum_{i=k+1}^{n} \frac{1-d_{\pi_\ell(i)}}{1-\xi}\varepsilon_{\pi_\ell(i)}(0).
\end{aligned}$$

Replacing $\varepsilon_{\pi_\ell(i)}(d)$ by $y_{\pi_\ell(i)}(d)$ would give $\hat{\tau}_{\mathtt{L}}^{\mathtt{IPW}}$ (or $\hat{\tau}_{\mathtt{R}}^{\mathtt{IPW}}$) instead of $\bar{\tau}_{\mathtt{L}}^{\mathtt{IPW}}$ (or $\bar{\tau}_{\mathtt{R}}^{\mathtt{IPW}}$). We use $\varepsilon_{\pi_\ell(i)}(d)$ to approximate $\mathscr{I}^{\mathtt{DIM}}(\cdot,\ell)$ because the constant arm means cancel from the DIM split contrast.

Under Definition SA-5, variance maximization for DIM chooses a splitting coordinate $\hat{\jmath}_{\mathtt{DIM}}$ and index $\hat{\imath}_{\mathtt{DIM}}$ whose child nodes

$$\mathtt{t}_{\mathtt{L}} = \{\mathbf{u} \in \mathcal{X} : u_{\hat{\jmath}_{\mathtt{DIM}}} \le x_{\pi_{\hat{\jmath}_{\mathtt{DIM}}}(\hat{\imath}_{\mathtt{DIM}}),\hat{\jmath}_{\mathtt{DIM}}}\} \qquad \text{and} \qquad \mathtt{t}_{\mathtt{R}} = \{\mathbf{u} \in \mathcal{X} : u_{\hat{\jmath}_{\mathtt{DIM}}} > x_{\pi_{\hat{\jmath}_{\mathtt{DIM}}}(\hat{\imath}_{\mathtt{DIM}}),\hat{\jmath}_{\mathtt{DIM}}}\},$$

maximize

$$\frac{n(\mathtt{t}_{\mathtt{L}})n(\mathtt{t}_{\mathtt{R}})}{n(\mathtt{t})}\left(\hat{\tau}_{\mathtt{DIM}}(\mathtt{t}_{\mathtt{L}}) - \hat{\tau}_{\mathtt{DIM}}(\mathtt{t}_{\mathtt{R}})\right)^2,$$

that is,

$$(\hat{\imath}_{\mathtt{DIM}}, \hat{\jmath}_{\mathtt{DIM}}) \in \operatorname*{arg\,max}_{(k,\ell)\in\mathcal{V}_{\mathtt{DIM}}} \mathscr{I}^{\mathtt{DIM}}(k,\ell).$$

If no valid root split exists, assign the auxiliary pair $(\hat{\imath}_{\mathtt{DIM}}, \hat{\jmath}_{\mathtt{DIM}}) = (1,1)$ and leave the root terminal. The probability of this event tends to zero, so it does not affect the limits below. We next control the random treatment-arm denominators on ranges with at least $r_n \to \infty$ observations in each child. The set $\mathcal{V}_{\mathtt{DIM}} = \mathcal{V}_{\mathtt{DIM}}(n)$ contains exactly the splits with positive treated and control denominators in both children. The approximation error below is maximized over these valid candidates, as required by Definition SA-5, with an empty maximum of nonnegative approximation errors defined as zero. On the balanced ranges, all required denominators are positive with probability tending to one.

Lemma SA-16 (Approximation Error). *Suppose Assumption SA-2 holds. Let $(r_n)_{n\in\mathbb{N}}$ be a sequence of real numbers such that $r_n \to \infty$. Then*

$$\max_{\substack{1\le\ell\le p,\ r_n\le k\le n-r_n \\ (k,\ell)\in\mathcal{V}_{\mathtt{DIM}}}} \left| \mathscr{I}^{\mathtt{DIM}}(k,\ell) - \bar{\mathscr{I}}^{\mathtt{IPW}}(k,\ell) \right| = O_{\mathbb{P}}\left(\sqrt{\frac{\log\log(n)}{r_n}}\right).$$

The following lemma complements this interior approximation and bounds the discarded endpoint scores.

Lemma SA-17 (Endpoint Scores). *Suppose Assumption SA-2 holds, and let $s_n \to \infty$ with $s_n = o(n)$. Then*

$$\max_{\substack{\ell\in[p],\ (k,\ell)\in\mathcal{V}_{\mathtt{DIM}} \\ 1\le k\le s_n \text{ or } n-s_n\le k<n}} \mathscr{I}^{\mathtt{DIM}}(k,\ell) \ + \max_{\substack{\ell\in[p] \\ 1\le k\le s_n \text{ or } n-s_n\le k<n}} \bar{\mathscr{I}}^{\mathtt{IPW}}(k,\ell) = O_{\mathbb{P}}(\log\log s_n).$$

An empty maximum of nonnegative scores is taken to be zero.

Rates for Decision Stumps. The preceding lemmas together transfer the joint block-maximum limits from $\bar{\mathscr{I}}^{\mathtt{IPW}}$ to $\mathscr{I}^{\mathtt{DIM}}$. Since the former is the CART criterion for the transformed outcome $\frac{d_i}{\xi}\varepsilon_i(1) - \frac{1-d_i}{1-\xi}\varepsilon_i(0)$, Section SA-2 gives the following split and error bounds.

Theorem SA-18 (Imbalanced Split)**.** *Suppose Assumption SA-2 holds. Then for each $a, b \in (0,1)$ with $a < b$, for every $\ell \in [p]$,*

$$\lim_{n\to\infty} \mathbb{P}\big(n^a \le \hat{\imath}_{\mathtt{DIM}} \le n^b,\ \hat{\jmath}_{\mathtt{DIM}} = \ell\big) = \frac{b-a}{2p} \qquad \text{and} \qquad \lim_{n\to\infty} \mathbb{P}\big(n - n^b \le \hat{\imath}_{\mathtt{DIM}} \le n - n^a,\ \hat{\jmath}_{\mathtt{DIM}} = \ell\big) = \frac{b-a}{2p}.$$

For each $\ell \in [p]$, $\mathbb{P}(\hat{\jmath}_{\mathtt{DIM}} = \ell) \to 1/p$. Conditional on $\hat{\jmath}_{\mathtt{DIM}} = \ell$, both window probabilities converge to $(b-a)/2$.

The imbalanced cells, in turn, yield the following uniform convergence rates.

Theorem SA-19 (Uniform Rates for Decision Stumps)**.** *Suppose Assumption SA-2 holds, and the tree has depth $K = 1$. Then $\sup_{\mathbf{x}\in\mathcal{X}} |\hat{\tau}_{\mathtt{DIM}}(\mathbf{x}) - \tau(\mathbf{x})| \to_{\mathbb{P}} 0$, and for any $a, b \in (0,1)$ with $a < b$,*

$$\lim_{n\to\infty} \mathbb{P}\left(\sup_{\mathbf{x}\in\mathcal{X}} |\hat{\tau}_{\mathtt{DIM}}(\mathbf{x}) - \tau(\mathbf{x})| \ge \sigma n^{-b/2}\sqrt{2\log\log(n)} \right) = b,$$

where $\sigma^2 = \mathbb{V}[\frac{d_i}{\xi}\varepsilon_i(1) - \frac{1-d_i}{1-\xi}\varepsilon_i(0)]$. Under the marginal probability integral transform normalization described after Assumption SA-2, for any deterministic sequence $\eta_n \downarrow 0$,

$$\liminf_{n\to\infty} \inf_{\mathbf{x}\in\mathcal{X}_n(a,\eta_n)} \mathbb{P}\Big(|\hat{\tau}_{\mathtt{DIM}}(\mathbf{x}) - \tau(\mathbf{x})| \ge \sigma n^{-b/2}\sqrt{(2+o(1))\log\log(n)} \Big) \ge \frac{b-a}{2p}.$$

Deeper Trees. The stump lower bound extends to recursive trees whose depth may vary with n, provided their leaf counts satisfy the condition below.

Theorem SA-20 (Uniform Rates for Deep Trees)**.** *Suppose Assumption SA-2 holds, and the NSS-DIM tree has depth at least one and at most a deterministic $K = K_n$. For every fixed $0 < a < b < 1$ satisfying*

$$2^K \log n = o\left(n^{a/2}\sqrt{\log\log n} \right),$$

we have

$$\liminf_{n\to\infty} \mathbb{P}\left(\sup_{\mathbf{x}\in\mathcal{X}} |\hat{\tau}_{\mathtt{DIM}}(\mathbf{x}) - \tau(\mathbf{x})| \ge \frac{\sigma}{2} n^{-b/2}\sqrt{\log\log n} \right) \ge b - a, \tag{SA-11}$$

where $\sigma^2 = \mathbb{V}[d_i\varepsilon_i(1)/\xi - (1-d_i)\varepsilon_i(0)/(1-\xi)]$. In particular, taking $a = b/2$ gives the probability lower bound $b/2$ under

$$2^K \log n = o\left(n^{b/4}\sqrt{\log\log n} \right).$$

Write

$$\rho_K = \limsup_{n\to\infty} \frac{K_n \log 2}{\log n}.$$

For every $b \in (0,1)$ with $2\rho_K < b$, we have

$$\liminf_{n\to\infty} \mathbb{P}\left(\sup_{\mathbf{x}\in\mathcal{X}} |\hat{\tau}_{\mathtt{DIM}}(\mathbf{x}) - \tau(\mathbf{x})| \ge \frac{\sigma}{2} n^{-b/2}\sqrt{\log\log n} \right) \ge b - 2\rho_K. \tag{SA-12}$$

For integrated L^q loss, $1 \le q < \infty$, we also have the following upper bound.

Theorem SA-21 (L^q Convergence Rate for NSS)**.** *Suppose Assumption SA-2 holds, and fix $1 \le q < \infty$. For $n \ge 2$ and any data-dependent axis-aligned tree with depth at most a deterministic $K = K_n$, the NSS-DIM terminal estimator satisfies*

$$\left(\mathbb{E}[\|\hat{\tau}_{\mathtt{DIM}} - \tau\|_q^q]\right)^{1/q} \le C_1 \left(\frac{2^K \log^{\max\{3,q+1\}}(n)}{n} \right)^{1/\max\{2,q\}}.$$

Moreover,

$$\mathbb{P}\left(\|\hat{\tau}_{\mathtt{DIM}} - \tau\|_q \ge C_2 \left(\frac{2^K \log^{\max\{3,q+1\}}(n)}{n} \right)^{1/\max\{2,q\}} \right) \le C_3 n^{-6}.$$

The positive constants C_1, C_2, and C_3 depend only on q, the fixed dimension p, and the distribution of $(y_i(0), y_i(1), d_i)$.

SA-3.2.2. Sample Splitting. With honest sample splitting, we obtain lower bounds on uniform error and upper bounds on integrated L^q error for every fixed $1 \le q < \infty$. The uniform-error lower bound, in this case, lacks the $\sqrt{\log\log n}$ factor present without sample splitting.

Theorem SA-22 (Uniform Rates for HON)**.** *Suppose Assumption SA-2 holds, and the HON-DIM tree splits its root whenever a valid split exists. Assume $\rho \le n_{\mathtt{T}}/n_\tau \le \rho^{-1}$ for some $\rho \in (0,1)$. Then for any $b \in (0,1)$,*

$$\liminf_{n\to\infty} \mathbb{P}\left(\sup_{\mathbf{x}\in\mathcal{X}} |\check{\tau}_{\mathtt{DIM}}(\mathbf{x}) - \tau(\mathbf{x})| \ge \frac{\sqrt{\rho(\sigma_0^2+\sigma_1^2)}}{2} n^{-b/2} \right) \ge \frac{\xi(1-\xi)(1-e^{-\rho/6})b}{8\max\{3, \mathbb{E}[\varepsilon_i(0)^4]/\sigma_0^4, \mathbb{E}[\varepsilon_i(1)^4]/\sigma_1^4\}},$$

where $\sigma_d^2 = \mathbb{V}[\varepsilon_i(d)]$ for $d \in \{0,1\}$.

Theorem SA-23 (L^q Convergence Rate for HON)**.** *Suppose Assumption SA-2 holds, and fix $1 \le q < \infty$. For any axis-aligned tree constructed using only $\mathcal{D}_{\mathtt{T}}$ and having depth at most a deterministic $K = K_n$, the honest DIM terminal estimator satisfies*

$$\left(\mathbb{E}[\|\check{\tau}_{\mathtt{DIM}} - \tau\|_q^q]\right)^{1/q} \le C_1 \left(\frac{2^K}{n_\tau+1}\right)^{1/\max\{2,q\}},$$

where $C_1, C_2 > 0$ depend only on q and the distribution of $(y_i(0), y_i(1), d_i)$, not on p or the fold sizes. Moreover, for every $\delta \in (0,1)$,

$$\mathbb{P}\left(\|\check{\tau}_{\mathtt{DIM}} - \tau\|_q \ge C_2 \left(\frac{2^K}{(n_\tau+1)\delta}\right)^{1/\max\{2,q\}} \right) \le \delta.$$

SA-3.3. SSE Estimator. The notation `SSE` refers to the splitting rule, with terminal treatment effects estimated by `DIM`. Thus `SSE` and `DIM` differ in how they construct the tree but use the same CATE estimator on a given partition. We now approximate the `SSE` criterion by a sum of transformed outcome regression criteria for treated and control observations. Two scalar strong approximations for the observed treatment-arm streams, followed by replacement of shared boundary observations, then give the limiting split distribution.

For a valid split of $\mathtt{t}$ into $\mathtt{t}_{\mathtt{L}}$ and $\mathtt{t}_{\mathtt{R}}$, profiling (SA-7) over the child intercepts and treatment coefficients shows that minimizing the residual sum of squares is equivalent to maximizing the sum of the treated and control variance gains. Terms independent of the candidate split drop out, leaving

$$\frac{n_1(\mathtt{t}_{\mathtt{L}})n_1(\mathtt{t}_{\mathtt{R}})}{n_1(\mathtt{t})}\left(\frac{1}{n_1(\mathtt{t}_{\mathtt{L}})}\sum_{i:\mathbf{x}_i\in\mathtt{t}_{\mathtt{L}}} d_i y_i - \frac{1}{n_1(\mathtt{t}_{\mathtt{R}})}\sum_{i:\mathbf{x}_i\in\mathtt{t}_{\mathtt{R}}} d_i y_i\right)^2 + \frac{n_0(\mathtt{t}_{\mathtt{L}})n_0(\mathtt{t}_{\mathtt{R}})}{n_0(\mathtt{t})}\left(\frac{1}{n_0(\mathtt{t}_{\mathtt{L}})}\sum_{i:\mathbf{x}_i\in\mathtt{t}_{\mathtt{L}}} (1-d_i) y_i - \frac{1}{n_0(\mathtt{t}_{\mathtt{R}})}\sum_{i:\mathbf{x}_i\in\mathtt{t}_{\mathtt{R}}} (1-d_i) y_i\right)^2.$$

Here

$$\hat{\mu}_{c,d} = \frac{1}{n_d(\mathtt{t}_c)}\sum_{i:\mathbf{x}_i\in\mathtt{t}_c} \mathbb{1}(d_i = d) y_i, \qquad c \in \{\mathtt{L},\mathtt{R}\}, \quad d \in \{0,1\}.$$

The map $(a,b) \mapsto (a, a+b)$ identifies the fitted control and treated means. Writing $\hat{\mu}_d = \{n_d(\mathtt{t}_{\mathtt{L}})\hat{\mu}_{\mathtt{L},d} + n_d(\mathtt{t}_{\mathtt{R}})\hat{\mu}_{\mathtt{R},d}\}/n_d(\mathtt{t})$,

$$\sum_{i:\mathbf{x}_i\in\mathtt{t},\ d_i=d} (y_i - \hat{\mu}_d)^2 = \sum_{c\in\{\mathtt{L},\mathtt{R}\}} \sum_{i:\mathbf{x}_i\in\mathtt{t}_c,\ d_i=d} (y_i - \hat{\mu}_{c,d})^2 + \frac{n_d(\mathtt{t}_{\mathtt{L}})n_d(\mathtt{t}_{\mathtt{R}})}{n_d(\mathtt{t})}(\hat{\mu}_{\mathtt{L},d} - \hat{\mu}_{\mathtt{R},d})^2.$$

Summing over $d = 0, 1$ gives the displayed criterion, so SSE chooses the split that maximizes the total improvement in the separate treated and control fits.

SA-3.3.1. No Sample Splitting.

Decision Stump. For a root split at rank k in coordinate ℓ, let $\mathtt{t}_{\mathtt{L}} = \{\mathbf{x} \in \mathcal{X} : x_\ell \le x_{\pi_\ell(k),\ell}\}$ and $\mathtt{t}_{\mathtt{R}} = \{\mathbf{x} \in \mathcal{X} : x_\ell > x_{\pi_\ell(k),\ell}\}$. The SSE objective, profiled over the child intercepts and treatment coefficients, is then

$$\min_{a_{\mathtt{L}},b_{\mathtt{L}},a_{\mathtt{R}},b_{\mathtt{R}}\in\mathbb{R}} \sum_{\mathbf{x}_i\in\mathtt{t}_{\mathtt{L}}} (y_i - a_{\mathtt{L}} - b_{\mathtt{L}} d_i)^2 + \sum_{\mathbf{x}_i\in\mathtt{t}_{\mathtt{R}}} (y_i - a_{\mathtt{R}} - b_{\mathtt{R}} d_i)^2 \qquad \text{[SA-13]}$$

over the valid index k and coordinate ℓ, with ties resolved by the prescribed rule. Denote the minimizers by $(\hat{\imath}_{\mathtt{SSE}}, \hat{\jmath}_{\mathtt{SSE}})$. The optimal threshold $\hat{\mathsf{s}}_{\mathtt{SSE}}$ in (SA-13) is $x_{\pi_{\hat{\jmath}_{\mathtt{SSE}}}(\hat{\imath}_{\mathtt{SSE}}),\hat{\jmath}_{\mathtt{SSE}}}$.

All ratios below are evaluated only at valid candidate splits under Definition SA-5. To rewrite (SA-13), let

$$\hat{\mu}_{\mathtt{L},0}(k,\ell) = \frac{\sum_{i=1}^{k}(1-d_{\pi_\ell(i)})y_{\pi_\ell(i)}}{\sum_{i=1}^{k}(1-d_{\pi_\ell(i)})}, \qquad \hat{\mu}_{\mathtt{L},1}(k,\ell) = \frac{\sum_{i=1}^{k} d_{\pi_\ell(i)} y_{\pi_\ell(i)}}{\sum_{i=1}^{k} d_{\pi_\ell(i)}},$$

$$\hat{\mu}_{\mathtt{R},0}(k,\ell) = \frac{\sum_{i=k+1}^{n}(1-d_{\pi_\ell(i)})y_{\pi_\ell(i)}}{\sum_{i=k+1}^{n}(1-d_{\pi_\ell(i)})}, \qquad \hat{\mu}_{\mathtt{R},1}(k,\ell) = \frac{\sum_{i=k+1}^{n} d_{\pi_\ell(i)} y_{\pi_\ell(i)}}{\sum_{i=k+1}^{n} d_{\pi_\ell(i)}}.$$

For $d \in \{0,1\}$, write

$$n_d = \sum_{i=1}^{n} \mathbb{1}(d_i = d), \qquad n_{\mathrm{L},d}(k,\ell) = \sum_{i=1}^{k} \mathbb{1}(d_{\pi_\ell(i)} = d), \qquad n_{\mathrm{R},d}(k,\ell) = n_d - n_{\mathrm{L},d}(k,\ell).$$

Suppressing (k,ℓ) in these counts, minimizing Equation (SA-13) is equivalent to maximizing

$$\mathscr{I}^{\mathrm{SSE}}(k,\ell) = \frac{n_{\mathrm{L},0} n_{\mathrm{R},0}}{n_0}(\hat{\mu}_{\mathrm{L},0}(k,\ell) - \hat{\mu}_{\mathrm{R},0}(k,\ell))^2 + \frac{n_{\mathrm{L},1} n_{\mathrm{R},1}}{n_1}(\hat{\mu}_{\mathrm{L},1}(k,\ell) - \hat{\mu}_{\mathrm{R},1}(k,\ell))^2.$$

The arm means μ_0 and μ_1 again cancel from the within-arm split contrasts. We therefore approximate the empirical criterion by the centered residual proxy

$$\mathscr{I}^{\mathrm{prox}}(k,\ell) = (1-\xi)\frac{k(n-k)}{n}(\bar{\mu}_{\mathrm{L},0}(k,\ell) - \bar{\mu}_{\mathrm{R},0}(k,\ell))^2 + \xi\frac{k(n-k)}{n}(\bar{\mu}_{\mathrm{L},1}(k,\ell) - \bar{\mu}_{\mathrm{R},1}(k,\ell))^2,$$

with

$$\bar{\mu}_{\mathrm{L},0}(k,\ell) = \frac{1}{k}\sum_{i=1}^{k} \frac{1-d_{\pi_\ell(i)}}{1-\xi}\varepsilon_{\pi_\ell(i)}(0), \qquad \bar{\mu}_{\mathrm{L},1}(k,\ell) = \frac{1}{k}\sum_{i=1}^{k} \frac{d_{\pi_\ell(i)}}{\xi}\varepsilon_{\pi_\ell(i)}(1),$$

$$\bar{\mu}_{\mathrm{R},0}(k,\ell) = \frac{1}{n-k}\sum_{i=k+1}^{n} \frac{1-d_{\pi_\ell(i)}}{1-\xi}\varepsilon_{\pi_\ell(i)}(0), \qquad \bar{\mu}_{\mathrm{R},1}(k,\ell) = \frac{1}{n-k}\sum_{i=k+1}^{n} \frac{d_{\pi_\ell(i)}}{\xi}\varepsilon_{\pi_\ell(i)}(1).$$

For a fixed coordinate $\ell \in [p]$, this proxy can, in turn, be approximated by the weighted squared norm of a standard bivariate time-transformed O-U process. Let $\mathcal{V}_{\mathrm{SSE}} = \mathcal{V}_{\mathrm{SSE}}(n)$ be the set of splits with positive treatment-arm denominators in both children. We maximize the approximation error over $\mathcal{V}_{\mathrm{SSE}}$, with an empty maximum of nonnegative approximation errors defined as zero.

Lemma SA-24 (Approximation Error)**.** *Suppose Assumption SA-2 holds. Let $(r_n)_{n\in\mathbb{N}}$ be a sequence of real numbers such that $r_n \to \infty$. Then*

$$\max_{\substack{1\le\ell\le p,\ r_n\le k\le n-r_n \\ (k,\ell)\in\mathcal{V}_{\mathrm{SSE}}}} \left|\mathscr{I}^{\mathrm{SSE}}(k,\ell) - \mathscr{I}^{\mathrm{prox}}(k,\ell)\right| = O_{\mathbb{P}}\left(\frac{1}{\sqrt{r_n}}\right).$$

Lemma SA-25 (Endpoint Scores)**.** *Suppose Assumption SA-2 holds, and let $s_n \to \infty$ with $s_n = o(n)$. Then*

$$\max_{\substack{\ell\in[p],\ (k,\ell)\in\mathcal{V}_{\mathrm{SSE}} \\ 1\le k\le s_n \text{ or } n-s_n\le k<n}} \mathscr{I}^{\mathrm{SSE}}(k,\ell) \ + \max_{\substack{\ell\in[p] \\ 1\le k\le s_n \text{ or } n-s_n\le k<n}} \mathscr{I}^{\mathrm{prox}}(k,\ell) = O_{\mathbb{P}}(\log\log s_n).$$

An empty maximum of nonnegative scores is taken to be zero.

Lemma SA-26 (SSE Bridge Coupling)**.** *Suppose Assumption SA-2 holds. Put $\sigma_d^2 = \mathbb{V}[\varepsilon_i(d)]$ and, for one fixed coordinate ℓ, define*

$$Z_{k,\ell} = \begin{pmatrix} \dfrac{\sqrt{(1-\xi)k(n-k)/n}}{\sigma_0}\left\{\bar{\mu}_{\mathrm{L},0}(k,\ell) - \bar{\mu}_{\mathrm{R},0}(k,\ell)\right\} \\ \dfrac{\sqrt{\xi k(n-k)/n}}{\sigma_1}\left\{\bar{\mu}_{\mathrm{L},1}(k,\ell) - \bar{\mu}_{\mathrm{R},1}(k,\ell)\right\} \end{pmatrix}.$$

For $1 \le r \le n/2$, this process can be coupled to two independent standard Brownian bridges B_0, B_1 so that

$$\max_{r\le k\le n-r} \left\| Z_{k,\ell} - \frac{(B_0(k/n), B_1(k/n))^\top}{\sqrt{(k/n)(1-k/n)}} \right\| = O_{\mathbb{P}}\left(\frac{\log n}{\sqrt{r}} + \frac{(\log n)^{3/4}}{r^{1/4}} + \sqrt{\frac{\log n}{n}}\right). \qquad \text{[SA-14]}$$

The two bridges are also independent of the treatment assignments. In particular, at $r_n = \lceil(\log n)^8\rceil$, the error is $o_{\mathbb{P}}((\log n)^{-1/2})$ and the corresponding uniform quadratic-score error is $o_{\mathbb{P}}(1)$.

Theorem SA-27 (Imbalanced Split for SSE)**.** *Suppose Assumption SA-2 holds. Then for each $a, b \in (0,1)$ with $a < b$, for every $\ell \in [p]$,*

$$\lim_{n\to\infty} \mathbb{P}\big(n^a \le \hat{\imath}_{\mathrm{SSE}} \le n^b,\ \hat{\jmath}_{\mathrm{SSE}} = \ell\big) = \frac{b-a}{2p} \qquad \text{and} \qquad \lim_{n\to\infty} \mathbb{P}\big(n - n^b \le \hat{\imath}_{\mathrm{SSE}} \le n - n^a,\ \hat{\jmath}_{\mathrm{SSE}} = \ell\big) = \frac{b-a}{2p}.$$

For each $\ell \in [p]$, $\mathbb{P}(\hat{\jmath}_{\mathrm{SSE}} = \ell) \to 1/p$. Conditional on $\hat{\jmath}_{\mathrm{SSE}} = \ell$, both window probabilities converge to $(b-a)/2$. Consequently,

$$\frac{\log\min\{\hat{\imath}_{\mathrm{SSE}}, n - \hat{\imath}_{\mathrm{SSE}}\}}{\log n}\ \Big|\ \hat{\jmath}_{\mathrm{SSE}} = \ell \ \rightsquigarrow\ \mathsf{Uniform}(0,1).$$

The same window and uniform limits hold without conditioning. These conclusions hold for any measurable tie rule. On $\{\mathcal{V}_{\mathrm{SSE}} = \varnothing\}$, assign the auxiliary rank and coordinate $(1,1)$ according to the root-split convention. The absence of a valid split has probability tending to zero and does not affect the limits.

Remark SA-1. Unequal treated and control residual variances give a weighted quadratic form in the bivariate O-U approximation. Lemma SA-41 then gives its extreme value limit. For equal variances, rescaling gives the squared Euclidean norm and Lemma SA-35 applies. ⌟

To pass from split locations to CATE error, we next control the residual contrast in the selected child.

Lemma SA-28 (Directional transfer for SSE stumps)**.** *Suppose Assumption SA-2 holds. For every $0 < a < b < 1$ and every $\ell \in [p]$, there exist positive constants C_1 and C_2, depending only on p and the distribution of $(y_i(0), y_i(1), d_i)$, such that, with $\mathtt{t}_{\mathtt{L}}$ denoting the selected left root child,*

$$\liminf_{n\to\infty} \mathbb{P}\left(\mathcal{V}_{\mathtt{SSE}} \neq \varnothing,\ n^a \le \hat{\imath}_{\mathtt{SSE}} \le n^b,\ \hat{\jmath}_{\mathtt{SSE}} = \ell,\ n(\mathtt{t}_{\mathtt{L}})\Delta(\mathtt{t}_{\mathtt{L}})^2 \ge C_1^2 \log\log n\right) \ge C_2(b-a).$$

The selected-child quantities are evaluated only on $\{\mathcal{V}_{\mathtt{SSE}} \neq \varnothing\}$. Substituting $\mathtt{t}_{\mathtt{R}}$ for $\mathtt{t}_{\mathtt{L}}$ gives the same conclusion on the right boundary event $n - n^b \le \hat{\imath}_{\mathtt{SSE}} \le n - n^a$. The displayed event therefore gives a depth one CATE error lower bound of order $n^{-b/2}\sqrt{\log\log n}$.

SSE and DIM use different splitting criteria but the same terminal CATE estimator on a fixed tree. Combining Theorem SA-27 and Lemma SA-28 with the DIM arguments gives the following results.

Theorem SA-29 (Uniform Rates for Decision Stumps)**.** *Suppose Assumption SA-2 holds and the tree has depth $K = 1$. For any $a, b \in (0,1)$ with $a < b$, we have*

$$\liminf_{n\to\infty} \mathbb{P}\left(\sup_{\mathbf{x}\in\mathcal{X}} |\hat{\tau}_{\mathtt{SSE}}(\mathbf{x}) - \tau(\mathbf{x})| \ge C_1 n^{-b/2}\sqrt{\log\log(n)}\right) \ge C_2 b,$$

where C_1 and C_2 are positive constants depending only on p and the distribution of $(y_i(0), y_i(1), d_i)$. Moreover, under the marginal probability integral transform normalization described after Assumption SA-2, for any deterministic sequence $\eta_n \downarrow 0$,

$$\liminf_{n\to\infty} \inf_{\mathbf{x}\in\mathcal{X}_n(a,\eta_n)} \mathbb{P}\left(|\hat{\tau}_{\mathtt{SSE}}(\mathbf{x}) - \tau(\mathbf{x})| \ge C_1 n^{-b/2}\sqrt{\log\log(n)}\right) \ge C_2(b-a).$$

Deeper Trees.

Theorem SA-30 (Uniform Rates for Deep Trees)**.** *Suppose Assumption SA-2 holds, and the NSS-SSE tree splits its root according to the prescribed criterion whenever a valid split exists and has depth at most a deterministic $K = K_n$, with $K \ge 1$. There exist positive constants $c_{\mathtt{SSE}}$ and $q_{\mathtt{SSE}}$, depending only on the distribution of $(y_i(0), y_i(1), d_i)$ and on the fixed covariate dimension p, such that, for every fixed $0 < a < b < 1$ satisfying*

$$2^K \log n = o\left(n^{a/2}\sqrt{\log\log n}\right),$$

we have

$$\liminf_{n\to\infty} \mathbb{P}\left(\sup_{\mathbf{x}\in\mathcal{X}} |\hat{\tau}_{\mathtt{SSE}}(\mathbf{x}) - \tau(\mathbf{x})| \ge c_{\mathtt{SSE}} n^{-b/2}\sqrt{\log\log n}\right) \ge 2q_{\mathtt{SSE}}(b-a). \tag{SA-15}$$

The constants do not depend on a, b, K, or n. In particular, taking $a = b/2$ gives the probability lower bound $q_{\mathtt{SSE}} b$ under

$$2^K \log n = o\left(n^{b/4}\sqrt{\log\log n}\right).$$

Write

$$\rho_K = \limsup_{n\to\infty} \frac{K_n \log 2}{\log n}.$$

For every $b \in (0,1)$ with $2\rho_K < b$, the same constants satisfy

$$\liminf_{n\to\infty} \mathbb{P}\left(\sup_{\mathbf{x}\in\mathcal{X}} |\hat{\tau}_{\mathtt{SSE}}(\mathbf{x}) - \tau(\mathbf{x})| \ge c_{\mathtt{SSE}} n^{-b/2}\sqrt{\log\log n}\right) \ge 2q_{\mathtt{SSE}}(b - 2\rho_K). \tag{SA-16}$$

Theorems SA-20 and SA-30 also remain valid when the prescribed root split is followed by any data-dependent refinement into at most 2^K axis-aligned cells containing both treatment arms, subject to their stated depth conditions.

Theorem SA-31 (L^q Convergence Rate for NSS)**.** *Suppose Assumption SA-2 holds, and fix $1 \le q < \infty$. For $n \ge 2$ and any NSS-SSE tree with depth at most a deterministic $K = K_n$, the NSS-SSE estimator satisfies*

$$\left(\mathbb{E}[\|\hat{\tau}_{\mathtt{SSE}} - \tau\|_q^q]\right)^{1/q} \le C_1 \left(\frac{2^K \log^{\max\{3,q+1\}}(n)}{n}\right)^{1/\max\{2,q\}}.$$

Moreover,

$$\mathbb{P}\left(\|\hat{\tau}_{\mathtt{SSE}} - \tau\|_q \ge C_2 \left(\frac{2^K \log^{\max\{3,q+1\}}(n)}{n}\right)^{1/\max\{2,q\}}\right) \le C_3 n^{-6}.$$

The positive constants C_1, C_2, and C_3 depend only on q, the fixed dimension p, and the distribution of $(y_i(0), y_i(1), d_i)$.

SA-3.3.2. Sample Splitting.

Theorem SA-32 (Uniform Rates for HON). *Suppose Assumption SA-2 holds. Assume $\rho \le n_{\mathsf{T}}/n_\tau \le \rho^{-1}$ for some $\rho \in (0,1)$. Then for any $b \in (0,1)$ and any HON-SSE tree that splits its root whenever a valid split exists,*

$$\liminf_{n\to\infty} \mathbb{P}\left(\sup_{\mathbf{x}\in\mathcal{X}} |\check{\tau}_{\mathtt{SSE}}(\mathbf{x}) - \tau(\mathbf{x})| \ge \frac{\sqrt{\rho(\sigma_0^2+\sigma_1^2)}}{2} n^{-b/2} \right) \ge \frac{\xi(1-\xi)(1-e^{-\rho/6})b}{8\max\{3, \mathbb{E}[\varepsilon_i(0)^4]/\sigma_0^4, \mathbb{E}[\varepsilon_i(1)^4]/\sigma_1^4\}},$$

where $\sigma_d^2 = \mathbb{V}[\varepsilon_i(d)]$ for $d \in \{0,1\}$.

Theorem SA-33 (L^q Convergence Rate for HON). *Suppose Assumption SA-2 holds, and fix $1 \le q < \infty$. For any HON-SSE tree constructed using only $\mathcal{D}_{\mathsf{T}}$ and having depth at most a deterministic $K = K_n$, the HON-SSE estimator satisfies*

$$\left(\mathbb{E}[\|\check{\tau}_{\mathtt{SSE}} - \tau\|_q^q]\right)^{1/q} \le C_1 \left(\frac{2^K}{n_\tau + 1}\right)^{1/\max\{2,q\}},$$

where $C_1, C_2 > 0$ depend only on q and the distribution of $(y_i(0), y_i(1), d_i)$, not on p or the fold sizes. Moreover, for every $\delta \in (0,1)$,

$$\mathbb{P}\left(\|\check{\tau}_{\mathtt{SSE}} - \tau\|_q \ge C_2 \left(\frac{2^K}{(n_\tau+1)\delta}\right)^{1/\max\{2,q\}} \right) \le \delta.$$

SA-3.4. Squared T-Statistic Estimators. Squared t-statistics also appear in causal-tree splitting rules [17]. Here we consider a pooled two-sample statistic applied to scalar pseudo-outcomes, for which the following equivalence holds. For $n \ge 3$, consider a scalar pseudo-outcome τ_i with left and right averages $\hat{\tau}_{\mathtt{L}}(k,\ell)$ and $\hat{\tau}_{\mathtt{R}}(k,\ell)$ after sorting along coordinate ℓ. The root split $(\hat{\imath}, \hat{\jmath})$ maximizes a criterion proportional to the squared pooled t-statistic over nondegenerate candidates,

$$(\hat{\imath}, \hat{\jmath}) \in \operatorname*{arg\,max}_{\substack{1\le k<n,\ \ell\in[p] \\ S^2(k,\ell)>0}} n \frac{(\hat{\tau}_{\mathtt{L}}(k,\ell) - \hat{\tau}_{\mathtt{R}}(k,\ell))^2}{S^2(k,\ell)/k + S^2(k,\ell)/(n-k)}.$$

The common factor n does not affect split selection. Here $S^2(k,\ell) > 0$ is the pooled within-child sample variance,

$$\begin{aligned} S^2(k,\ell) &= \frac{1}{n-2}\sum_{i=1}^{k} (\tau_{\pi_\ell(i)} - \hat{\tau}_{\mathtt{L}}(k,\ell))^2 + \frac{1}{n-2}\sum_{i=k+1}^{n} (\tau_{\pi_\ell(i)} - \hat{\tau}_{\mathtt{R}}(k,\ell))^2 \\ &= \frac{1}{n-2}\left[\sum_{i=1}^{n} (\tau_i - n^{-1}\sum_{j=1}^{n}\tau_j)^2 - \frac{k(n-k)}{n}(\hat{\tau}_{\mathtt{L}}(k,\ell) - \hat{\tau}_{\mathtt{R}}(k,\ell))^2 \right]. \end{aligned}$$

Let

$$A(k,\ell) = \frac{k(n-k)}{n}(\hat{\tau}_{\mathtt{L}}(k,\ell) - \hat{\tau}_{\mathtt{R}}(k,\ell))^2, \qquad T_0 = \sum_{i=1}^{n} (\tau_i - n^{-1}\sum_{j=1}^{n}\tau_j)^2.$$

Then, for candidates with $T_0 > A(k,\ell)$,

$$n\frac{(\hat{\tau}_{\mathtt{L}}(k,\ell) - \hat{\tau}_{\mathtt{R}}(k,\ell))^2}{S^2(k,\ell)/k + S^2(k,\ell)/(n-k)} = \frac{n(n-2)A(k,\ell)}{T_0 - A(k,\ell)}.$$

For $0 \le A < T_0$,

$$\frac{\mathrm{d}}{\mathrm{d}A}\frac{n(n-2)A}{T_0 - A} = \frac{n(n-2)T_0}{(T_0-A)^2} > 0.$$

Thus the two criteria select the same split among candidates with positive pooled variance. If no such candidate exists, choose a fixed deterministic valid split, or leave the node unsplit if none is valid. If the pseudo-outcomes τ_i satisfy Assumption SA-1 in place of y_i, the strong law for the pseudo-outcomes and their squares [18, Theorem 2.4.1], together with (SA-26), gives

$$\frac{T_0}{n} \to \mathbb{V}[\tau_i] > 0 \quad \text{almost surely}, \qquad \frac{\max_{\substack{1\le k<n \\ 1\le \ell\le p}} A(k,\ell)}{T_0} = O_{\mathbb{P}}\left(\frac{\log\log n}{n}\right) = o_{\mathbb{P}}(1).$$

Hence

$$\mathbb{P}\big(T_0 > A(k,\ell) \text{ for all } k,\ell\big) \to 1,$$

and the fallback rule does not affect the lower bounds.

The equivalence applies to the pooled statistic for scalar pseudo-outcomes displayed above. Studentizing the treatment and control means separately, using heteroskedastic or sandwich standard errors, or changing the candidate-specific denominator in another way can, however, change the ordering relative to variance maximization.

SA-3.5. Unbiasedness.

Lemma SA-34 (Unbiasedness)**.** *Suppose Assumption SA-2 holds. In each formula below, let* $\mathtt{t}_{\mathtt{HON}}(\mathbf{x})$ *denote the terminal node used by the estimator on the left-hand side for honest estimation at* $\mathbf{x}$*. Then*

$$\mathbb{E}[\check{\tau}_{\mathtt{IPW}}(\mathbf{x})] = \tau(\mathbf{x}) - \tau(\mathbf{x})\mathbb{P}(n(\mathtt{t}_{\mathtt{HON}}(\mathbf{x})) = 0),$$
$$\mathbb{E}[\check{\tau}_l(\mathbf{x})] = \tau(\mathbf{x}) - \tau(\mathbf{x})\mathbb{P}(n_0(\mathtt{t}_{\mathtt{HON}}(\mathbf{x})) = 0 \text{ or } n_1(\mathtt{t}_{\mathtt{HON}}(\mathbf{x})) = 0), \qquad l \in \{\mathtt{DIM}, \mathtt{SSE}\}.$$

For the NSS statements, use Definition SA-5, allowing early stopping and pruning that replaces a subtree by the root of that subtree. With covariates and treatment assignments fixed, let $\mathcal{P}_l(\mathbf{z})$ *denote the final partition for a residual vector* $\mathbf{z} = (z_1, \ldots, z_n)^\top$*. Here the responses are* $y_i = \mu_{d_i} + z_i$ *for* $l \in \{\mathtt{DIM}, \mathtt{SSE}\}$*, and the transformed outcomes satisfy* $y_i(d_i - \xi)/\{\xi(1-\xi)\} = \tau + z_i$ *for* $l = \mathtt{IPW}$*. Assume*

$$\mathcal{P}_l(\mathbf{z}) = \mathcal{P}_l(-\mathbf{z}), \qquad \mathbf{z} \in \mathbb{R}^n, \quad l \in \{\mathtt{IPW}, \mathtt{DIM}, \mathtt{SSE}\},$$

where the final partition includes all stopping, pruning, and tie-breaking decisions. For NSS-DIM and NSS-SSE, suppose additionally that each marginal potential error distribution is symmetric,

$$\varepsilon_i(d) \overset{d}{=} -\varepsilon_i(d), \qquad d = 0, 1.$$

Then, for $l \in \{\mathtt{DIM}, \mathtt{SSE}\}$*,*

$$\mathbb{E}[\hat{\tau}_l(\mathbf{x})] = \tau(\mathbf{x})\big\{1 - \xi^n - (1-\xi)^n\big\}.$$

For $n \geq 2$*,*

$$\mathbb{E}\left[\hat{\tau}_l(\mathbf{x}) \,\middle|\, 0 < \sum_{i=1}^n d_i < n\right] = \tau(\mathbf{x}).$$

For NSS-IPW, the symmetry condition is that, with

$$\tilde{\varepsilon}_i = y_i \frac{d_i - \xi}{\xi(1-\xi)} - \tau,$$

the transformed residual vector $(\tilde{\varepsilon}_1, \ldots, \tilde{\varepsilon}_n)^\top$ *is centrally symmetric conditional on the treatment assignments. Alternatively, it suffices that* $\tilde{\varepsilon}_i \overset{d}{=} -\tilde{\varepsilon}_i$ *unconditionally, provided the final IPW partition, including stopping and pruning, depends only on the covariates and transformed outcomes. Under either IPW symmetry condition,*

$$\mathbb{E}[\check{\tau}_{\mathtt{IPW}}(\mathbf{x})] = \tau(\mathbf{x}).$$

All these NSS statements hold for arbitrary τ *under the respective conditions.*

The partition condition permits stopping and pruning based on covariates, total node counts, and invariant split scores, with invariant tie rules. No deterministic depth bound is required.

Remark SA-2 (Sufficient conditions for IPW symmetry)**.** To verify the transformed residual symmetry in Lemma SA-34, suppose both marginal potential error distributions are symmetric. If

$$(1-\xi)\mu_1 + \xi\mu_0 = 0,$$

then, conditional on the treatment assignments,

$$\tilde{\varepsilon}_i = \begin{cases} \varepsilon_i(1)/\xi, & d_i = 1, \\ -\varepsilon_i(0)/(1-\xi), & d_i = 0, \end{cases}$$

and the transformed residual vector is centrally symmetric. In the balanced experiment case $\xi = 1/2$, this centering condition reduces to $\mu_1 + \mu_0 = 0$. The centering condition allows any τ, since it holds when $\mu_1 = \xi\tau$ and $\mu_0 = -(1-\xi)\tau$.

The unconditional residual symmetry in Lemma SA-34 also holds when $\xi = 1/2$ and $\varepsilon_i(1) \overset{d}{=} \varepsilon_i(0)$. Indeed,

$$\mathcal{L}(\tilde{\varepsilon}_i) = \tfrac{1}{2}\mathcal{L}\big(\mu_1 + \mu_0 + 2\varepsilon_i(1)\big) + \tfrac{1}{2}\mathcal{L}\big(-\mu_1 - \mu_0 - 2\varepsilon_i(0)\big) = \mathcal{L}(-\tilde{\varepsilon}_i).$$

This holds for arbitrary μ_0, μ_1 and does not require symmetry of the common marginal error distribution. ⌟

SA-4. Correction to Eicker (1979)

[1, Theorem 5, p. 120, Eqs. (1.17)–(1.18)] considers scalar standardized partial-sum maxima over indices $n^{o(1)} < k < n^c$, with $0 < c < 1$. The stated one-sided limit is $\exp\{-c^2 e^{-t}/(2\sqrt{\pi})\}$, whose square is the stated limit for the absolute maximum. The factor c^2 should instead be c, so the shift $2\log c$ in the limiting distribution should be $\log c$. The limit for the absolute maximum is still the square of the one-sided limit. With $t = z - \frac{1}{2}\log\pi$, the corrected absolute-maximum limit becomes $\exp(-ce^{-z})$.

For the associated Gaussian process, logarithmic time gives an interval of length $(c + o(1))\log n$. Rescaling [3, Lemma 2.1] then gives the corrected vector O-U limit below. The Gaussian correlation inequality [19, Theorem 2 and Remark 3(i)] also excludes the factor c^2. For a centered Gaussian process U, two adjacent windows I_1, I_2, and any threshold v, the symmetric convex events

$$A_h = \left\{ \sup_{t \in I_h} \|U(t)\| \le v \right\}, \qquad h = 1, 2,$$

satisfy $\mathbb{P}(A_1 \cap A_2) \ge \mathbb{P}(A_1)\mathbb{P}(A_2)$. If $|I_h| = c_h \log n$, with $c_1, c_2 > 0$ and $c_1 + c_2 < 1$, the proposed c^2 limit would require

$$e^{-(c_1+c_2)^2 e^{-z}} \ge e^{-(c_1^2+c_2^2)e^{-z}}, \qquad (c_1 + c_2)^2 > c_1^2 + c_2^2,$$

which is impossible. The corrected limit instead satisfies

$$e^{-(c_1+c_2)e^{-z}} = e^{-c_1 e^{-z}} e^{-c_2 e^{-z}}.$$

Lemma SA-35 (Vector Valued Markov Type Darling–Erdős)**.** *Let $\{(V_1(t), \dots, V_d(t)) : 0 \le t < \infty\}$ be d independent identically distributed Ornstein–Uhlenbeck processes with $\mathbb{E}[V_i(t)] = 0$ and $\mathbb{E}[V_i(t)V_i(s)] = \exp(-|t-s|/2)$, $1 \le i \le d$. Define*

$$N(t) = \left(\sum_{1 \le i \le d} V_i^2(t) \right)^{1/2}.$$

For any $c > 0$, $z \in \mathbb{R}$,

$$\lim_{n\to\infty} \mathbb{P}\Big(a(\log(n)) \sup_{0 \le t \le c\log(n)} N(t) - b_d(\log(n)) \le z \Big) = \exp\Big(- e^{-(z - \log(c))} \Big),$$

where $a(t) = (2\log(t))^{1/2}$ and $b_d(t) = 2\log(t) + \frac{d}{2}\log\log(t) - \log\Gamma(d/2)$. Here $\Gamma(\cdot)$ denotes the Euler gamma function. The same limit holds, uniformly for z in compact sets, over $[s_n, s_n + c_n \log n]$ for deterministic $s_n \ge 0$ and $c_n \to c > 0$, with strict or weak inequality and deterministic $o(1)$ perturbations of z.

SA-4.1. Proof of Lemma SA-35.

Put

$$M_n(c) = \sup_{0 \le t \le c\log n} N(t), \qquad L = \log\log n, \qquad A = \log c.$$

By [3, Lemma 2.1, pp. 673–674],

$$a(c\log n) M_n(c) - b_d(c\log n) \rightsquigarrow G, \qquad \mathbb{P}(G \le z) = e^{-e^{-z}}.$$

The change of normalization is then

$$\begin{aligned} \frac{a(\log n)}{a(c\log n)} &= \left(1 + \frac{A}{L}\right)^{-1/2} = 1 - \frac{A}{2L} + O(L^{-2}), \\ b_d(c\log n) - b_d(\log n) &= 2A + \frac{d}{2}\log\left(1 + \frac{A}{L}\right) = 2A + O(L^{-1}), \\ \frac{a(\log n)}{a(c\log n)} b_d(c\log n) - b_d(\log n) &= A + O\left(\frac{1 + \log L}{L}\right). \end{aligned}$$

Consequently,

$$\begin{aligned} a(\log n) M_n(c) - b_d(\log n) = {} & \frac{a(\log n)}{a(c\log n)} \{ a(c\log n) M_n(c) - b_d(c\log n) \} \\ & + \frac{a(\log n)}{a(c\log n)} b_d(c\log n) - b_d(\log n) \rightsquigarrow G + \log c. \end{aligned}$$

Slutsky's theorem [18, Exercises 3.2.12–3.2.14, pp. 124–125] therefore gives

$$\mathbb{P}\big(a(\log n) M_n(c) - b_d(\log n) \le z \big) \to e^{-ce^{-z}}.$$

Stationarity also covers deterministic translates. For $c_n \to c > 0$ and $0 < \eta < c$, eventually

$$M_n(c-\eta) \le M_n(c_n) \le M_n(c+\eta).$$

Letting $\eta \downarrow 0$ proves the same limit for lengths $(c+o(1))\log n$. Monotonicity in z and continuity of the limiting CDF give uniformity for z in compact sets and permit strict inequalities and deterministic $o(1)$ changes in z.

The scalar correction also holds under the hypotheses of [1, Theorem 2.1, pp. 121–122], without Gaussianity. Let Y_i be independent with $\mathbb{E}[Y_i] = 0$, $\mathbb{E}[Y_i^2] = 1$, and $\sup_{i\ge1} \mathbb{E}[|Y_i|^3] < \infty$, and put

$$Z_k = \sum_{i=1}^{k} Y_i, \qquad S_m = \max_{1\le k\le m} \frac{Z_k}{\sqrt{k}}, \qquad S_m^{\mathrm{abs}} = \max_{1\le k\le m} \frac{|Z_k|}{\sqrt{k}}.$$

In Eicker's notation, write

$$a_n = \sqrt{2\log\log n}, \qquad b_n = 2\log\log n + \tfrac{1}{2}\log\log\log n, \qquad D(t) = \exp\{-e^{-t}/(2\sqrt{\pi})\}.$$

His Theorem 2.1 gives limiting CDFs D and D^2 for $a_m S_m - b_m$ and $a_m S_m^{\mathrm{abs}} - b_m$, respectively. For $m = \lfloor n^c \rfloor$, $0 < c < 1$, and $L = \log\log n$,

$$\frac{a_n}{a_m} = 1 - \frac{\log c}{2L} + o(L^{-1}), \qquad b_m = b_n + 2\log c + o(1), \qquad b_n = 2L + O(\log L).$$

Consequently,

$$a_n S_m - b_n = \frac{a_n}{a_m}(a_m S_m - b_m) + \frac{a_n}{a_m} b_m - b_n = a_m S_m - b_m + \log c + o_{\mathbb{P}}(1).$$

The same identity holds with S_m^{abs} in place of S_m. Their limiting CDFs under the n normalization are therefore $D(t-\log c)$ and $D(t-\log c)^2$. For a positive null sequence ϵ_n and any fixed $0 < \eta < c$, eventually $\epsilon_n < \eta$, so

$$\limsup_{n\to\infty} \mathbb{P}\left(\max_{1\le k\le n^{\epsilon_n}} \frac{|Z_k|}{\sqrt{k}} \ge \frac{b_n + t}{a_n} \right) \le 1 - D(t-\log\eta)^2 \to 0 \qquad (\eta \downarrow 0).$$

The lower range can therefore be deleted from both maxima without changing their limits. Since $\mathbb{E}[Z_k^2/k] = 1$ and $(b_n+t)/a_n \to \infty$, changing an integer endpoint also has vanishing effect. The corrected limits thus hold on Eicker's stated range $n^{o(1)} < k < n^c$.

SA-5. Proofs

SA-5.1. Technical Lemmas. A coupling for an array can be transferred to any array with the same law by drawing the coupled process from its conditional law given that array. For independent arrays, these draws use mutually independent auxiliary randomness, independent of all original variables, so the resulting array–process pairs are independent.

We use one-coordinate strong approximations to obtain the results for any fixed number of covariates. Throughout the proofs, unless otherwise specified, take

$$L_n = \log\log n, \qquad r_n = \lceil (\log n)^8 \rceil, \qquad m_n = \lfloor n/(\log n)^4 \rfloor.$$

All assertions concern sufficiently large n, so $1 < r_n < m_n < n/2$.

Moment and maximal inequalities. We use Jensen's, Hölder's, and Markov's inequalities as stated in [18, Theorems 1.6.2–1.6.4, pp. 29–30]. Cauchy–Schwarz is the case of Hölder's inequality with both exponents equal to two. Applying Markov's inequality to $(X - \mathbb{E}[X])^2$ similarly gives Chebyshev's inequality. If $(M_k)_{k=0}^m$ is a scalar or finite-dimensional vector martingale and $r > 1$, Doob's maximal inequality [18, Theorem 4.4.4, p. 236] gives

$$\mathbb{E}\left[\max_{0\le k\le m} \|M_k\|^r \right] \le \left(\frac{r}{r-1}\right)^r \mathbb{E}[\|M_m\|^r]. \qquad \text{[SA-17]}$$

For scalars, the norm is absolute value. Put $M^* = \max_{0\le k\le m} \|M_k\|$ and take $t, R > 0$. Optional sampling for the nonnegative submartingale $\|M_k\|$, integration, and Hölder's inequality give

$$\begin{aligned} t\mathbb{P}(M^* \ge t) &\le \mathbb{E}[\|M_m\| \mathbb{1}(M^* \ge t)], \\ \mathbb{E}[(M^* \wedge R)^r] &\le \frac{r}{r-1}\mathbb{E}[\|M_m\|(M^* \wedge R)^{r-1}] \le \frac{r}{r-1}\{\mathbb{E}[\|M_m\|^r]\}^{1/r}\{\mathbb{E}[(M^* \wedge R)^r]\}^{(r-1)/r}. \end{aligned}$$

Divide and let $R \to \infty$ to obtain (SA-17).

For independent centered scalar increments X_i with uniformly bounded second and fourth moments, expansion of the fourth power gives

$$\mathbb{E}\left[\left(\sum_{i=1}^{m} X_i\right)^4\right] = \sum_{i=1}^{m} \mathbb{E}[X_i^4] + 6 \sum_{1\le i<j\le m} \mathbb{E}[X_i^2]\mathbb{E}[X_j^2] \le Cm^2.$$

Thus, by (SA-17),

$$\mathbb{E}\left[\max_{1\le k\le m}\left|\sum_{i=1}^{k} X_i\right|^2\right] \le Cm \qquad \text{and} \qquad \mathbb{E}\left[\max_{1\le k\le m}\left|\sum_{i=1}^{k} X_i\right|^4\right] \le Cm^2.$$

Cauchy–Schwarz also bounds products of dependent partial-sum maxima.

Exponential concentration. The exponential moment assumption gives, for some $h > 0$,

$$\mathbb{E}[e^{h|\varepsilon_i|}] \le \mathbb{E}[e^{h\varepsilon_i}] + \mathbb{E}[e^{-h\varepsilon_i}] < \infty, \qquad \mathbb{E}[|\varepsilon_i|^r] < \infty \quad (r > 0).$$

For centered X with this property, Taylor's formula and dominated convergence [18, Theorem 1.6.7, p. 30] give fixed $b, C > 0$ such that

$$\log \mathbb{E}[e^{\lambda X}] \le C\lambda^2 \quad (|\lambda| \le b).$$

For independent copies $X_1, \dots, X_m$, exponential Markov's inequality then gives

$$\mathbb{P}\left(\left|\sum_{i=1}^{m} X_i\right| > x\right) \le 2 \inf_{0\le\lambda\le b} \exp(-\lambda x + Cm\lambda^2) \le 2\exp\left\{-c\min\left\{\frac{x^2}{m}, x\right\}\right\}.$$

Equivalently, after enlarging C,

$$\mathbb{P}\left(\left|\sum_{i=1}^{m} X_i\right| > C\{\sqrt{mt} + t\}\right) \le 2e^{-t}, \qquad t > 0. \tag{SA-18}$$

This is the fixed-law form of Bernstein's inequality in [20, Proposition 2.7.1 and Theorem 2.8.1]. It also applies to either of the finitely many centered arm variables used below, after enlarging C. For $B \sim \mathsf{Binomial}(m, u)$, the same exponential Markov argument with $\lambda = \log 2$ gives

$$\mathbb{P}(B \le mu/2) \le e^{\lambda mu/2}(1 - u + ue^{-\lambda})^m \le e^{-(1-\log 2)mu/2} \le e^{-mu/8}. \tag{SA-19}$$

We also use these inequalities conditionally, specifying at each application why the conditioning preserves the independence and laws of the variables.

Brownian maximal bounds. For a standard d-dimensional Brownian motion W, the reflection principle [21, Section 6.1, pp. 38–39], the Gaussian tail bound [18, Theorem 1.2.6, p. 13], and a union bound give

$$\mathbb{P}\left(\sup_{0\le t\le h} \|W(t)\| > x\right) \le \sum_{j=1}^{d} \mathbb{P}\left(\sup_{0\le t\le h} |W_j(t)| > x/\sqrt{d}\right) \le 4d\exp\{-x^2/(2dh)\}, \qquad h, x > 0. \tag{SA-20}$$

The same bound holds for increments on any deterministic interval of length h. Thus a union bound over M such intervals introduces a factor M, without requiring independence between them.

Lemma SA-36 (Joint O-U block maxima)**.** *Let $U = (U(t))_{t\in\mathbb{R}}$ be a stationary d-dimensional standard O-U process with covariance $\mathbb{E}[U(s)U(t)^\top] = e^{-|t-s|/2} I_d$, and let $q : \mathbb{R}^d \to \mathbb{R}$ be continuous. Suppose that, for $w_n(u) = \beta_n + Au$, where $A > 0$, and some $\kappa > 0$,*

$$\mathbb{P}\left(\sup_{0\le t\le c\log n} q(U(t)) \le w_n(u)\right) \to \exp(-c\kappa e^{-u}) \qquad (c > 0,\ u \in \mathbb{R}).$$

Let $I_{1,n}, \dots, I_{s,n}$ be a fixed finite collection of deterministic intervals, in time order, with disjoint interiors and lengths $(c_h + o(1))\log n$, where $c_h > 0$. Then

$$\left(\frac{\sup_{t\in I_{h,n}} q(U(t)) - \beta_n}{A}\right)_{h=1}^{s} \rightsquigarrow (G_h)_{h=1}^{s}, \qquad \mathbb{P}(G_h \le u) = \exp(-c_h\kappa e^{-u}),$$

where the G_h are independent. Any deterministic interval of length $o(\log n)$ has maximum exceeding $w_n(u)$, or attaining $w_n(u)$, with probability tending to zero. Finally,

$$\mathbb{P}\left(G_h > \max_{\substack{j\in[s]\\ j\ne h}} G_j\right) = \frac{c_h}{\sum_{j=1}^{s} c_j}.$$

Proof. For the mixing bound, put

$$\mathscr{F}_s^- = \sigma\{U(t) : t \le s\}, \qquad \mathscr{F}_{s+h}^+ = \sigma\{U(t) : t \ge s+h\}.$$

For $E \in \mathscr{F}_s^-$ and $F \in \mathscr{F}_{s+h}^+$, let $f(\mathbf{x}) = \mathbb{P}(E \mid U(s) = \mathbf{x})$ and $g(\mathbf{y}) = \mathbb{P}(F \mid U(s+h) = \mathbf{y})$. With P_h the O-U transition operator, the Markov property [21, Section 4.4.3, p. 18] gives

$$\mathbb{P}(E \cap F) = \mathbb{E}[\mathbb{1}(E)(P_h g)(U(s))] = \mathbb{E}[f(U(s))(P_h g)(U(s))] = \mathbb{E}[f(U(s))g(U(s+h))],$$

where $0 \le f, g \le 1$. The joint law P_ρ of these two Gaussian vectors has correlation $\rho = e^{-h/2}$, whereas the product law P_0 has zero correlation. Let $L = \mathrm{d}P_\rho/\mathrm{d}P_0$. With $\Sigma_\rho = \left(\begin{smallmatrix} I_d & \rho I_d \\ \rho I_d & I_d \end{smallmatrix}\right)$, Gaussian integration gives

$$\mathbb{E}_{P_0}[L^2] = \det(\Sigma_\rho)^{-1} \det(2\Sigma_\rho^{-1} - I_{2d})^{-1/2} = (1-\rho^2)^{-d} \qquad \text{and} \qquad \det(2\Sigma_\rho^{-1} - I_{2d}) = \left(\frac{1-\rho}{1+\rho}\frac{1+\rho}{1-\rho}\right)^d = 1.$$

Since $0 \le fg \le 1$, total variation and Cauchy–Schwarz give

$$|\mathbb{P}(E \cap F) - \mathbb{P}(E)\mathbb{P}(F)| \le \frac{1}{2}\mathbb{E}_{P_0}[|L-1|] \le \frac{1}{2}\{\mathbb{E}_{P_0}[(L-1)^2]\}^{1/2} = \frac{1}{2}\{(1-e^{-h})^{-d} - 1\}^{1/2} = O(e^{-h/2}).$$

Stationarity and a squeeze between lengths $(c \pm \eta)\log n$ extend the assumed marginal law to intervals of length $(c + o(1))\log n$. Similarly, an interval J_n of length $o(\log n)$ is eventually shorter than $\eta \log n$, so

$$\limsup_{n\to\infty} \mathbb{P}\left(\sup_{t \in J_n} q(U(t)) > w_n(u)\right) \le 1 - \exp(-\eta\kappa e^{-u}).$$

Let $\eta \downarrow 0$. Replacing u by $u - \epsilon$ also proves the assertion with $\ge$.

Write $I_{h,n} = [a_{h,n}, b_{h,n}]$ and set

$$\lambda_n = \sqrt{\log n}, \qquad I'_{h,n} = [a_{h,n}, b_{h,n} - \lambda_n], \qquad E_{h,n} = \left\{\sup_{t \in I'_{h,n}} q(U(t)) \le w_n(u_h)\right\}.$$

These intervals retain coefficients c_h and have gaps at least λ_n. Since

$$\bigcap_{1 \le j < h} E_{j,n} \in \mathscr{F}^-_{a_{h,n} - \lambda_n}, \qquad E_{h,n} \in \mathscr{F}^+_{a_{h,n}},$$

iteration of the mixing bound gives

$$\left|\mathbb{P}\left(\bigcap_{h=1}^s E_{h,n}\right) - \prod_{h=1}^s \mathbb{P}(E_{h,n})\right| \le C(s-1)e^{-\lambda_n/2} = o(1).$$

The removed pieces, in turn, satisfy

$$\sum_{h=1}^s \mathbb{P}\left(\sup_{t \in I_{h,n} \setminus I'_{h,n}} q(U(t)) > w_n(u_h)\right) = o(1).$$

Restoring them then proves the joint limit. Continuity gives the same suprema on half-open intervals and their closures.

The variables e^{-G_h} are independent exponentials with rates $c_h\kappa$, because $\mathbb{P}(e^{-G_h} > t) = \mathbb{P}(G_h < -\log t) = e^{-c_h\kappa t}$ for $t > 0$. Consequently,

$$\mathbb{P}\left(G_h > \max_{\substack{j \in [s] \\ j \ne h}} G_j\right) = \int_0^\infty c_h \kappa \exp\left\{-t\kappa \sum_{j=1}^s c_j\right\} \mathrm{d}t = \frac{c_h}{\sum_{j=1}^s c_j}.$$

The last assertion follows, and the limiting block maxima have no ties. □

Lemma SA-37 (Transfer of block maxima)**.** *Suppose two arrays of scores differ uniformly by $o_{\mathbb{P}}(1)$ on a retained candidate set, partitioned into a fixed finite number of deterministic blocks. If the normalized block maxima of the first array, centered by β_n and divided by a fixed $A > 0$, converge jointly to a vector with no ties for its largest component, then the second array has the same joint limit and the same limiting winning-block probabilities. The conclusion holds for every measurable rule resolving empirical ties. Candidates outside the retained set can be included if their probability of attaining the retained maximum tends to zero.*

Proof. Let $\mathcal{J}_{1,n},\dots,\mathcal{J}_{s,n}$ be the disjoint blocks of the retained set, and write J_n and $\widetilde{J}_n$ for the two score arrays. Define

$$M_{h,n}=\max_{v\in\mathcal{J}_{h,n}}J_n(v),\qquad \widetilde{M}_{h,n}=\max_{v\in\mathcal{J}_{h,n}}\widetilde{J}_n(v),\qquad \Delta_n=\max_{v\in\bigcup_{h=1}^{s}\mathcal{J}_{h,n}}|J_n(v)-\widetilde{J}_n(v)|.$$

Let $G=(G_1,\dots,G_s)$ denote the assumed limit and let $\widehat{h}_n$ be the block selected from the second array. Slutsky's theorem then gives

$$\max_{1\le h\le s}|M_{h,n}-\widetilde{M}_{h,n}|\le\Delta_n=o_{\mathbb{P}}(1),\qquad \left(\frac{\widetilde{M}_{h,n}-\beta_n}{A}\right)_{h=1}^{s}\rightsquigarrow G.$$

Every tie rule satisfies

$$\left\{M_{h,n}>\max_{\substack{j\in[s]\\ j\neq h}}M_{j,n}+2\Delta_n\right\}\subseteq\{\widehat{h}_n=h\}\subseteq\left\{M_{h,n}\ge\max_{\substack{j\in[s]\\ j\neq h}}M_{j,n}-2\Delta_n\right\}.$$

For $\epsilon>0$, the portmanteau theorem [18, Theorem 3.10.1, pp. 196–197] and $\mathbb{P}(2\Delta_n>A\epsilon)\to0$ imply

$$\mathbb{P}\left(G_h>\max_{\substack{j\in[s]\\ j\neq h}}G_j+\epsilon\right)\le\liminf_{n\to\infty}\mathbb{P}(\widehat{h}_n=h)\le\limsup_{n\to\infty}\mathbb{P}(\widehat{h}_n=h)\le\mathbb{P}\left(G_h\ge\max_{\substack{j\in[s]\\ j\neq h}}G_j-\epsilon\right).$$

Let $\epsilon\downarrow0$. The absence of a tie for the limiting largest component therefore proves the claimed winning probability. Adding discarded candidates, finally, changes it by at most their probability of attaining the retained maximum. □

Lemma SA-38 (Replacement of shared boundary observations)**.** *Let $V_1,\dots,V_n$ be i.i.d. centered vectors in a fixed-dimensional Euclidean space, with finite covariance Σ, and let $\pi_1,\dots,\pi_p$ be independent uniform permutations, independent of the vectors, with fixed p. Form the first m_n and reversed last m_n labels of each permutation, giving $2p$ arrays $(i_{a,h}:1\le h\le m_n)$. There are mutually independent i.i.d. arrays $(\widetilde{V}_{a,h})$ with the law of V_1 such that, for every fixed quadratic form Q,*

$$\max_{1\le a\le2p}\max_{1\le k\le m_n}\left|Q\left(\frac{\sum_{h=1}^{k}V_{i_{a,h}}-(k/n)\sum_{i=1}^{n}V_i}{\sqrt{k(1-k/n)}}\right)-Q\left(\frac{\sum_{h=1}^{k}\widetilde{V}_{a,h}}{\sqrt{k}}\right)\right|=O_{\mathbb{P}}((\log n)^{-1}).$$

The resulting arrays are also independent of all the permutations.

Proof. Order the arrays deterministically and define the positions to be replaced by

$$\mathcal{R}_a=\left\{1\le h\le m_n:i_{a,h}\in\bigcup_{1\le b<a}\{i_{b,t}:1\le t\le m_n\}\right\}.$$

Let $V'_{a,h}$ be independent copies of V_1, independent of all original observations and permutations, and set

$$\widetilde{V}_{a,h}=\begin{cases}V_{i_{a,h}}, & h\notin\mathcal{R}_a,\\ V'_{a,h}, & h\in\mathcal{R}_a.\end{cases}$$

Given the permutations, all entries use distinct original observations or independent copies. Thus

$$\mathcal{L}\big((\widetilde{V}_{a,h})_{a,h}\mid\pi_1,\dots,\pi_p\big)=\bigotimes_{a=1}^{2p}\bigotimes_{h=1}^{m_n}\mathcal{L}(V_1),$$

which proves both independence assertions.

For a fixed array let $D_a(k)=\sum_{h=1}^{k}(V_{i_{a,h}}-\widetilde{V}_{a,h})$ and $R_a(k)=\sum_{h=1}^{k}\mathbb{1}(h\in\mathcal{R}_a)$. Conditional on the permutations, the increments of D_a are independent and centered, because no label occurs twice within an array. Each nonzero increment, more specifically, is the difference of two independent copies and has conditional squared norm expectation $2\,\mathrm{tr}(\Sigma)$. Thus (SA-17), with $r=2$, gives

$$\mathbb{E}\left[\max_{1\le k\le t}\|D_a(k)\|^2\,\middle|\,\pi_1,\dots,\pi_p\right]\le8\,\mathrm{tr}(\Sigma)R_a(t).$$

The two arrays from one permutation are disjoint. An array from another permutation, by contrast, contains any given label with probability m_n/n, so $\mathbb{E}[R_a(t)]\le(2p-2)tm_n/n$. For a dyadic range $2^j\le k<2^{j+1}$,

$$\mathbb{E}\left[\max_{\substack{2^j\le k<2^{j+1}\\ k\le m_n}}\frac{\|D_a(k)\|^2}{k}\right]\le2^{-j}\mathbb{E}\left[\max_{1\le k\le\min\{2^{j+1},m_n\}}\|D_a(k)\|^2\right]\le C_p\frac{m_n}{n}.$$

Sum over the at most $1+\log_2 m_n$ ranges and the fixed number $2p$ of arrays. This gives

$$\mathbb{E}\left[\max_{1\le a\le 2p}\max_{1\le k\le m_n}\frac{\|D_a(k)\|^2}{k}\right]\le C_p\frac{m_n}{n}(1+\log m_n).$$

The same dyadic argument for the original and replacement partial sums gives

$$\max_{1\le a\le 2p}\max_{1\le k\le m_n}\frac{\|\sum_{h=1}^k V_{i_{a,h}}\|+\|\sum_{h=1}^k \widetilde{V}_{a,h}\|}{\sqrt{k}}=O_{\mathbb{P}}(\sqrt{\log n}).$$

Also $\|\sum_{i=1}^n V_i\|/\sqrt{n}=O_{\mathbb{P}}(1)$. For the bridge correction, put $S_a(k)=\sum_{h=1}^k V_{i_{a,h}}$ and $S(n)=\sum_{i=1}^n V_i$. Then

$$\frac{S_a(k)-(k/n)S(n)}{\sqrt{k(1-k/n)}}-\frac{S_a(k)}{\sqrt{k}}=\left\{(1-k/n)^{-1/2}-1\right\}\frac{S_a(k)}{\sqrt{k}}-\sqrt{\frac{k}{n}}(1-k/n)^{-1/2}\frac{S(n)}{\sqrt{n}}.$$

Since $k\le m_n<n/2$, its norm is uniformly bounded by

$$O_{\mathbb{P}}\left(\frac{m_n}{n}\sqrt{\log n}+\sqrt{\frac{m_n}{n}}\right).$$

Write $H_a(k)$ and $\widetilde{H}_a(k)$ for the two vectors in the statement. The preceding bounds together give

$$\max_{\substack{1\le a\le 2p\\1\le k\le m_n}}\|H_a(k)-\widetilde{H}_a(k)\|=O_{\mathbb{P}}\left(\sqrt{\frac{m_n}{n}\log n}\right)\qquad\text{and}\qquad\max_{\substack{1\le a\le 2p\\1\le k\le m_n}}\{\|H_a(k)\|+\|\widetilde{H}_a(k)\|\}=O_{\mathbb{P}}(\sqrt{\log n}).$$

For $Q(\mathbf{v})=\mathbf{v}^\top\mathbf{C}\mathbf{v}$, take $\mathbf{C}$ symmetric without loss of generality. Then

$$|Q(\mathbf{v})-Q(\mathbf{w})|\le\|\mathbf{C}\|\|\mathbf{v}-\mathbf{w}\|(\|\mathbf{v}\|+\|\mathbf{w}\|)\qquad\text{and}\qquad\max_{\substack{1\le a\le 2p\\1\le k\le m_n}}|Q(H_a(k))-Q(\widetilde{H}_a(k))|=O_{\mathbb{P}}\left(\sqrt{\frac{m_n}{n}}\log n\right)=O_{\mathbb{P}}((\log n)^{-1}).$$

Reversing a right boundary changes the bridge's sign and leaves Q unchanged. □

Lemma SA-39 (Extension from one coordinate to fixed dimension)**.** *In the setting of Lemma SA-38, let Q be a nonnegative quadratic form and write*

$$H_n(k,j)=\frac{\sum_{i=1}^k V_{\pi_j(i)}-(k/n)\sum_{i=1}^n V_i}{\sqrt{k(1-k/n)}}.$$

Suppose that for one i.i.d. ordering there is a standard vector Brownian bridge B, possibly depending on n, such that

$$\max_{r_n\le k\le n-r_n}\left|Q(H_n(k,1))-Q\left(\frac{\Sigma^{1/2}B(k/n)}{\sqrt{(k/n)(1-k/n)}}\right)\right|=o_{\mathbb{P}}(1).\qquad\text{[SA-21]}$$

Suppose the stationary score $q(U)=Q(\Sigma^{1/2}U)$ satisfies the marginal hypothesis of Lemma SA-36 with $w_n(u)=\beta_n+Au$ and $\kappa>0$. Let $\mathcal{V}_n$ be the set of valid candidates and put

$$\mathcal{R}_n=\{(k,j):r_n\le k\le n-r_n,\ 1\le j\le p\}.$$

Let $J_n(k,j)$ be actual scores, defined arbitrarily on invalid candidates, such that $\mathbb{P}(\mathcal{R}_n\subseteq\mathcal{V}_n)\to 1$ and

$$\max_{1\le j\le p}\max_{r_n\le k\le n-r_n}|J_n(k,j)-Q(H_n(k,j))|=o_{\mathbb{P}}(1).$$

All maxima of actual scores are over valid candidates, with an empty maximum assigned $-\infty$. Assume also that, for each fixed u,

$$\mathbb{P}\left(\max_{(k,j)\in\mathcal{V}_n\setminus\mathcal{R}_n}J_n(k,j)\ge w_n(u)\right)\to 0.$$

Then the $2p$ boundaries admit joint independent Gumbel block limits under any fixed finite subdivision by powers n^a, $0<a<1$. Each boundary has length coefficient one, and each subinterval of boundary distances $[n^a,n^b]$ has coefficient $b-a$. Finite unions of these blocks have the sum of their coefficients. In particular, for a fixed coordinate ℓ and $0<a<b<1$, take the target set

$$\{(k,\ell):n^a\le k\le n^b\}.$$

Its maximum, the complementary maximum in coordinate ℓ, and the full maxima in each other coordinate have independent normalized limits with coefficients

$$b-a, \qquad 2-b+a, \qquad 2,\dots,2.$$

Here normalization means subtraction of β_n and division by A, and a coefficient c gives CDF $\exp(-c\kappa e^{-u})$. Every measurable maximizing rule therefore selects the target with probability tending to $(b-a)/(2p)$, and

$$\mathbb{P}\left(\max_{(k,j)\in\mathcal{V}_n} J_n(k,j) \le w_n(u)\right) \to \exp(-2p\kappa e^{-u}). \tag{SA-22}$$

The same conclusions hold for reflected target ranks. If $\beta_n/L_n \to q_\star > 0$, then $\max_{(k,j)\in\mathcal{V}_n} J_n(k,j)/(q_\star L_n) \to_{\mathbb{P}} 1$.

Proof. For $0 < s \le t < 1$, each coordinate B_ν of a standard Brownian bridge satisfies

$$\frac{\mathbb{E}[B_\nu(s)B_\nu(t)]}{\sqrt{s(1-s)t(1-t)}} = \sqrt{\frac{s(1-t)}{t(1-s)}} = \exp\left\{-\frac{1}{2}\left|\log\frac{t}{1-t} - \log\frac{s}{1-s}\right|\right\}.$$

Thus the centered Gaussian processes satisfy

$$\left\{\frac{B(t)}{\sqrt{t(1-t)}} : 0<t<1\right\} \overset{d}{=} \left\{U\left(\log\frac{t}{1-t}\right) : 0<t<1\right\}.$$

Adjacent retained ranks have logit spacing at most C/r_n. From the O-U transition formula, with W a standard Brownian motion of the same dimension,

$$U(s+h) - U(s) = (e^{-h/2}-1)U(s) + e^{-h/2}\int_0^h e^{v/2}\,\mathrm{d}W(s+v).$$

The integral process is Brownian motion at time $e^h - 1$. For $0 \le h \le \delta \le 1$,

$$|e^{-h/2}-1| \le C\delta, \qquad e^h - 1 \le C\delta.$$

Gaussian tails, (SA-20), and a union bound over the grid points $s_i = \log\{i/(n-i)\}$, $r_n \le i \le n - r_n$, yield, in turn,

$$\mathbb{P}\left(\max_{r_n\le i\le n-r_n}\sup_{0\le h\le\delta}\|U(s_i+h)-U(s_i)\| > C(\delta+\sqrt{\delta})x\right) \le Cne^{-cx^2}, \qquad x \ge 1.$$

Take $\delta = C/r_n$ and $x = C\sqrt{\log n}$ with C sufficiently large. Then

$$\max_{r_n\le i\le n-r_n}\sup_{0\le h\le\delta}\|U(s_i+h)-U(s_i)\| = O_{\mathbb{P}}\left(\sqrt{\frac{\log n}{r_n}}\right),$$

$$\max_{r_n\le i\le n-r_n}\sup_{0\le h\le\delta}\|U(s_i+h)\| = O_{\mathbb{P}}(\sqrt{\log n}),$$

$$\max_{r_n\le i\le n-r_n}\sup_{0\le h\le\delta}|q(U(s_i+h))-q(U(s_i))| = O_{\mathbb{P}}\left(\frac{\log n}{\sqrt{r_n}}\right) = o_{\mathbb{P}}(1).$$

The last line uses that q is quadratic. Fixed-power subdivision points and their adjacent cells lie in the retained range eventually. Lemma SA-36 and (SA-21) therefore give the one-coordinate joint block limits.

Apply Lemma SA-38 to pass from one coordinate to the full collection of boundaries. Extend each of its $2p$ independent i.i.d. arrays independently to length n, and apply the one-coordinate coupling separately to each extension. The product of these couplings therefore gives independent Gaussian bridges for the independent extensions. Write

$$\widetilde{S}_a(k) = \sum_{h=1}^{k}\widetilde{V}_{a,h}, \qquad \widetilde{H}_a(k) = \frac{\widetilde{S}_a(k) - (k/n)\widetilde{S}_a(n)}{\sqrt{k(1-k/n)}}.$$

The maximal bounds in that lemma give

$$\max_{1\le a\le 2p}\max_{1\le k\le m_n}\left\|\widetilde{H}_a(k) - \frac{\widetilde{S}_a(k)}{\sqrt{k}}\right\| = O_{\mathbb{P}}\left(\frac{m_n}{n}\sqrt{\log n} + \sqrt{\frac{m_n}{n}}\right),$$

$$\max_{1\le a\le 2p}\max_{1\le k\le m_n}\left|Q(\widetilde{H}_a(k)) - Q\left(\frac{\widetilde{S}_a(k)}{\sqrt{k}}\right)\right| = O_{\mathbb{P}}\left(\frac{m_n}{n}\log n + \sqrt{\frac{m_n}{n}\log n}\right) = o_{\mathbb{P}}(1).$$

The bridges are independent across the $2p$ arrays. Their logit lengths are

$$\log\frac{m_n}{n-m_n}-\log\frac{r_n}{n-r_n}=\log n+o(\log n),$$
$$\log\frac{n^b}{n-n^b}-\log\frac{n^a}{n-n^a}=(b-a)\log n+o(\log n),\qquad 0<a<b<1.$$

Lemma SA-36, interpolation, replacement, and the actual-score approximation together give the claimed joint boundary laws. Endpoint rounding changes only intervals of length $o(\log n)$ and therefore does not change the limits.

For the middle range,

$$2\log\frac{n-m_n}{m_n}=O(\log\log n)=o(\log n).$$

The one-coordinate coupling, the short-interval bound, and a finite union bound give, in turn,

$$\mathbb{P}\left(\max_{\substack{1\le j\le p\\ m_n<k<n-m_n}} J_n(k,j)\ge w_n(u)\right)\to 0.$$

The assumed endpoint bound treats the remaining discarded candidates. For independent block limits and any nonempty set H of block indices, coefficients add because

$$\mathbb{P}\left(\max_{h\in H}G_h\le u\right)=\prod_{h\in H}e^{-c_h\kappa e^{-u}}=e^{-(\sum_{h\in H}c_h)\kappa e^{-u}}.$$

This gives the target/complement limits and (SA-22).

Define the retained boundary set

$$\mathcal{B}_n=\{(k,j):r_n\le k\le m_n \text{ or } n-m_n\le k\le n-r_n,\ 1\le j\le p\},$$

and let $B_n=\max_{(k,j)\in\mathcal{V}_n\cap\mathcal{B}_n}J_n(k,j)$ and $O_n=\max_{(k,j)\in\mathcal{V}_n\setminus\mathcal{B}_n}J_n(k,j)$. For every fixed u,

$$\mathbb{P}(O_n\ge B_n)\le\mathbb{P}(O_n\ge w_n(u))+\mathbb{P}(B_n\le w_n(u)).$$

The first term tends to zero and the second to $\exp(-2p\kappa e^{-u})$. Let $u\to-\infty$ to see that the probability of an outside candidate attaining the boundary maximum tends to zero. Lemma SA-37 and the exponential calculation in Lemma SA-36 then give the claimed winning probabilities for every measurable tie rule. The reflected target has the same length coefficient, so the tie rule need not be reflection invariant. Tightness of the normalized full maximum proves the final assertion. □

Lemma SA-40 (Dyadic product bounds)**.** *Fix $d\in\{0,1\}$. In one covariate ordering, let*

$$A_k=\sum_{i=1}^{k}\mathbb{1}(d_i=d)\varepsilon_i(d),\qquad N_k=\sum_{i=1}^{k}\mathbb{1}(d_i=d),\qquad B_k=N_k-\xi_d k.$$

Under Assumption SA-2, for $r\to\infty$,

$$\max_{r\le k\le n}\frac{|A_kB_k|}{k^{3/2}}=O_{\mathbb{P}}(r^{-1/2}),\qquad \max_{r\le k\le n}\frac{|A_k|^2|B_k|}{k^2}=O_{\mathbb{P}}(r^{-1/2}),$$
$$\mathbb{P}\left(\exists k\in[r,n]:\ N_k<\xi_dk/2\right)=O(r^{-1}).$$

The conclusions hold in reverse order and simultaneously over the fixed set of covariate coordinates. Maxima over empty ranges are zero.

Proof. The ordered pairs are i.i.d. because the ordering is independent of assignments and errors. Hence A_k and B_k are centered martingales. By (SA-17), the fourth-moment expansion, and Cauchy–Schwarz, for every integer $K\ge1$,

$$\mathbb{E}\left[\max_{1\le k\le 2K}|A_kB_k|\right]\le\left(\mathbb{E}\left[\max_{1\le k\le 2K}|A_k|^2\right]\right)^{1/2}\left(\mathbb{E}\left[\max_{1\le k\le 2K}|B_k|^2\right]\right)^{1/2}\le CK,$$
$$\mathbb{E}\left[\max_{1\le k\le 2K}|A_k|^2|B_k|\right]\le\left(\mathbb{E}\left[\max_{1\le k\le 2K}|A_k|^4\right]\right)^{1/2}\left(\mathbb{E}\left[\max_{1\le k\le 2K}|B_k|^2\right]\right)^{1/2}\le CK^{3/2},$$
$$\mathbb{E}\left[\max_{1\le k\le 2K}|B_k|^2\right]\le CK.$$

Extend the sequence independently beyond n if needed and set $K_j = 2^j\lceil r\rceil$. Then

$$\mathbb{E}\left[\max_{r\le k\le n}\frac{|A_kB_k|}{k^{3/2}}\right] \le \sum_{j\ge 0} K_j^{-3/2}\mathbb{E}\left[\max_{1\le k\le 2K_j}|A_kB_k|\right] \le C\sum_{j\ge 0}K_j^{-1/2} \le Cr^{-1/2},$$

$$\mathbb{E}\left[\max_{r\le k\le n}\frac{|A_k|^2|B_k|}{k^2}\right] \le \sum_{j\ge 0} K_j^{-2}\mathbb{E}\left[\max_{1\le k\le 2K_j}|A_k|^2|B_k|\right] \le Cr^{-1/2},$$

$$\mathbb{P}\left(\exists k\in[r,n] : N_k < \xi_d k/2\right) \le \sum_{j\ge 0}\mathbb{P}\left(\max_{1\le k\le 2K_j}|B_k| > \xi_d K_j/2\right) \le C\sum_{j\ge 0}K_j^{-1} \le C/r.$$

Markov's inequality then gives the two stochastic orders. Reversal and a finite union bound finally give the remaining assertions. □

Lemma SA-41 (Weighted bivariate O-U maxima)**.** *Let U_0 and U_1 be independent standard O-U processes with covariance $\mathbb{E}[U_d(s)U_d(t)] = \exp(-|s-t|/2)$, and let*

$$Q_\Lambda(t) = \lambda_0 U_0(t)^2 + \lambda_1 U_1(t)^2, \qquad \Lambda = (\lambda_0,\lambda_1),$$

where $\lambda_0,\lambda_1 > 0$. There exist thresholds $w_n(u;\Lambda)$ and a constant $\kappa_\Lambda \in (0,\infty)$, both independent of c, such that, for every fixed $c > 0$, uniformly for u in compact subsets of $\mathbb{R}$,

$$\mathbb{P}\left(\sup_{0\le t\le c\log n} Q_\Lambda(t) \le w_n(u;\Lambda)\right) = \exp\{-c\kappa_\Lambda e^{-u}\} + o(1),$$

where, with $\lambda_\star = \max\{\lambda_0,\lambda_1\}$ and $\beta_n(\Lambda) = w_n(0;\Lambda)$,

$$w_n(u;\Lambda) = \beta_n(\Lambda) + 2\lambda_\star u = 2\lambda_\star\log\log n + O(\log\log\log n + |u|).$$

If $\lambda_0 = \lambda_1$, this reduces to Lemma SA-35 after rescaling.

Proof. When $\lambda_0 = \lambda_1$, set

$$w_n(u;\Lambda) = 2\lambda_0\{\log\log n + \log\log\log n + u\}, \qquad \kappa_\Lambda = 1.$$

For bounded u,

$$a(\log n)\sqrt{w_n(u;\Lambda)/\lambda_0} - b_2(\log n) = u + o(1),$$

so Lemma SA-35 gives the result.

Suppose now that $\lambda_1 > \lambda_0$ and put $U = (U_0,U_1)^\top$. With the coordinates ordered as (U_1,U_0), the hypotheses of [4, Theorem 1, p. 126, conditions (4)–(6), pp. 125–126] hold with

$$D = \mathrm{diag}(\lambda_1,\lambda_0), \qquad \alpha = 1, \qquad C_0 = I_2/2, \qquad c_{11} = 1/2,$$
$$\mathbb{E}[U(0)U(t)^\top] = e^{-|t|/2}I_2 = I_2 - |t|I_2/2 + o(|t|), \qquad I_2 \pm e^{-|t|/2}I_2 \succ 0 \quad (t>0).$$

For fixed $h > 0$ and $w\to\infty$, with Pickands constant H_1, its asymptotic expression for $\mathbb{P}(\sup_{0\le t\le h} Q_\Lambda(t) > w)$ is

$$\frac{1}{2\lambda_1}\frac{2}{\sqrt{1-\lambda_0/\lambda_1}}hH_1 w\frac{\sqrt{\lambda_1}}{\sqrt{2\pi w}}e^{-w/(2\lambda_1)}.$$

The Pickands constant in this normalization is $H_1 = 1$. To verify the normalization directly, let W be standard Brownian motion and put $M_T = \sup_{0\le t\le T}\{\sqrt{2}W(t) - t\}$. With Φ and ϕ denoting the standard normal CDF and density, respectively, the reflection principle and the constant-drift change of measure [21, Sections 5.6.1 and 6.1] give, for $x \ge 0$,

$$\mathbb{P}(M_T > x) = 1 - \Phi\left(\frac{x+T}{\sqrt{2T}}\right) + e^{-x}\left\{1 - \Phi\left(\frac{x-T}{\sqrt{2T}}\right)\right\}.$$

Put $s = \sqrt{2T}$ and $a = T/s$. Integration by parts and $e^x\phi((x+T)/s) = \phi((x-T)/s)$ give

$$\int_0^\infty e^x\{1-\Phi((x+T)/s)\}\,\mathrm{d}x = -\{1-\Phi(a)\} + \int_0^\infty \phi((x-T)/s)\,\frac{\mathrm{d}x}{s} = 2\Phi(a) - 1,$$
$$\int_0^\infty \{1-\Phi((x-T)/s)\}\,\mathrm{d}x = T\Phi(a) + s\phi(a).$$

Therefore

$$\mathbb{E}[e^{M_T}] = 1 + \int_0^\infty e^x \mathbb{P}(M_T > x)\,\mathrm{d}x = (T+2)\Phi(a) + s\phi(a).$$

Consequently $H_1 = \lim_{T\to\infty} T^{-1}\mathbb{E}[e^{M_T}] = 1$. Substituting above then gives, for each fixed $h > 0$,

$$\mathbb{P}\left(\sup_{0\le t\le h} Q_\Lambda(t) > w\right) \sim h\, r(w), \qquad r(w) = \frac{\sqrt{w}}{\sqrt{2\pi(\lambda_1-\lambda_0)}} \exp\!\Big(-\frac{w}{2\lambda_1}\Big), \qquad w \to \infty.$$

The same expansion holds with $\ge$ in place of $>$, by squeezing between thresholds w and $w - w^{-1}$.

To obtain the limit over growing intervals, define

$$D_w = \{\mathbf{x} \in \mathbb{R}^2 : \lambda_0 x_0^2 + \lambda_1 x_1^2 < w\}, \qquad \tau_w = \inf\{t \ge 0 : U(t) \notin D_w\},$$

where $U = (U_0, U_1)^\top$, and kill the process at τ_w. Let $(\theta_{w,j}, f_{w,j})_{j\ge1}$ be the orthonormal Dirichlet eigenbasis of $-(\Delta - \mathbf{x}\cdot\nabla)/2$ on D_w in $L^2(\gamma_2)$, ordered by increasing eigenvalue, and put

$$(\theta_w, f_w) = (\theta_{w,1}, f_{w,1}), \qquad a_{w,j} = \int_{D_w} f_{w,j}\,\mathrm{d}\gamma_2.$$

Here γ_2 is standard Gaussian measure. The bounded smooth ellipse has a discrete self-adjoint spectrum [5, Section 2.1, p. 2190, Eqs. (2.8)–(2.9)]. That source's operator is, however, twice the negative generator used here. Extend the eigenfunctions by zero outside D_w.

The unrestricted O-U semigroup satisfies

$$P_t f(\mathbf{x}) = \int f\big(e^{-t/2}\mathbf{x} + \sqrt{1-e^{-t}}\,\mathbf{y}\big)\,\mathrm{d}\gamma_2(\mathbf{y}), \qquad \nabla P_t f = e^{-t/2} P_t \nabla f.$$

For $f \in C_c^\infty(\mathbb{R}^2)$, the displayed formula and dominated convergence give $P_t f \to \int f\,\mathrm{d}\gamma_2$ in $L^2(\gamma_2)$ as $t \to \infty$. Gaussian integration by parts, followed by Jensen's inequality and stationarity, gives

$$\mathbb{V}_{\gamma_2}[f] = -\int_0^\infty \frac{\mathrm{d}}{\mathrm{d}t}\|P_t f\|^2_{L^2(\gamma_2)}\,\mathrm{d}t = \int_0^\infty \|\nabla P_t f\|^2_{L^2(\gamma_2)}\,\mathrm{d}t \le \int_0^\infty e^{-t}\,\mathrm{d}t \int \|\nabla f\|^2\,\mathrm{d}\gamma_2.$$

Approximate $f_w \in H_0^1(D_w, \gamma_2)$ by smooth compactly supported functions and pass to the limit after zero extension. The eigenfunction identity then gives

$$1 - a_{w,1}^2 = \mathbb{V}_{\gamma_2}[f_w] \le \int \|\nabla f_w\|^2\,\mathrm{d}\gamma_2 = 2\theta_w.$$

Write $\mathrm{d}U(t) = -\frac{1}{2}U(t)\,\mathrm{d}t + \mathrm{d}\mathbf{W}(t)$, where $\mathbf{W}$ is a standard two-dimensional Brownian motion. For each eigenfunction, Itô's formula [21, Section 5.3, pp. 25–26], localized to inner ellipses, with passage to the boundary justified by the regularity and gradient bound established below, gives

$$e^{\theta_{w,j}(T\wedge\tau_w)} f_{w,j}(U(T\wedge\tau_w)) = f_{w,j}(\mathbf{x}) + \int_0^{T\wedge\tau_w} e^{\theta_{w,j}s} \nabla f_{w,j}(U(s))\cdot \mathrm{d}\mathbf{W}(s).$$

The eigenfunction satisfies

$$-\nabla\cdot\Big(e^{-\|\mathbf{x}\|^2/2}\nabla f_{w,j}(\mathbf{x})\Big) = 2\theta_{w,j} e^{-\|\mathbf{x}\|^2/2} f_{w,j}(\mathbf{x}).$$

On the bounded ellipse,

$$0 < e^{-w/(2\min\{\lambda_0,\lambda_1\})} \le e^{-\|\mathbf{x}\|^2/2} \le 1,$$

so weighted and ordinary Sobolev norms are equivalent. The coefficients are also smooth and uniformly elliptic. The eigenfunction belongs to $H_0^1(D_w)$, hence to $L^2(D_w)$. Elliptic boundary regularity then gives

$$f_{w,j} \in H^k(D_w) \implies 2\theta_{w,j} e^{-\|\cdot\|^2/2} f_{w,j} \in H^k(D_w) \implies f_{w,j} \in H^{k+2}(D_w), \qquad k = 0, 1, \ldots.$$

Iteration and Sobolev embedding [22, Theorem 4.31 and Corollary 4.32, p. 116] give $f_{w,j} \in C^\infty(\overline{D}_w)$. Its zero Sobolev trace is therefore a pointwise zero boundary value. Hence

$$\mathbb{E}_{\mathbf{x}}\left[\int_0^{T\wedge\tau_w} e^{2\theta_{w,j}s}\|\nabla f_{w,j}(U(s))\|^2\,\mathrm{d}s\right] \le T e^{2\theta_{w,j}T}\|\nabla f_{w,j}\|^2_{\infty,\overline{D}_w} < \infty.$$

The stopped stochastic integral therefore has mean zero, so

$$\mathbb{E}_{\mathbf{x}}[f_{w,j}(U(T))\mathbb{1}(\tau_w > T)] = e^{-T\theta_{w,j}} f_{w,j}(\mathbf{x}).$$

For the killed semigroup $P_T^{D_w} g(\mathbf{x}) = \mathbb{E}_{\mathbf{x}}[g(U(T))\mathbb{1}(\tau_w > T)]$, Jensen's inequality and stationarity give, after zero extension of g,

$$\|P_T^{D_w} g\|_{L^2(\gamma_2)}^2 \le \int \mathbb{E}_{\mathbf{x}}[g(U(T))^2 \mathbb{1}(\tau_w > T)]\, \mathrm{d}\gamma_2(\mathbf{x}) \le \|g\|_{L^2(\gamma_2)}^2.$$

Thus the eigenfunction expansion converges in $L^2(\gamma_2)$, and

$$\mathbb{P}_{\gamma_2}(\tau_w > T) = \langle \mathbb{1}(D_w), P_T^{D_w}\mathbb{1}(D_w)\rangle_{L^2(\gamma_2)} = \sum_{j\ge 1} a_{w,j}^2 e^{-T\theta_{w,j}}, \qquad \sum_{j\ge 1} a_{w,j}^2 = \gamma_2(D_w) \le 1.$$

All summands are nonnegative, $\theta_{w,j} \ge \theta_w$, and $a_{w,1}^2 \ge 1 - 2\theta_w$. Since paths are continuous, $\{\tau_w > T\} = \{\sup_{0\le t\le T} Q_\Lambda(t) < w\}$. Consequently, for every $T > 0$,

$$(1 - 2\theta_w)e^{-T\theta_w} \le \mathbb{P}\left(\sup_{0\le t\le T} Q_\Lambda(t) < w\right) \le e^{-T\theta_w}.$$

For fixed $h > 0$, the middle probability tends to one, so $\theta_w \to 0$. The same bounds also imply

$$1 - e^{-h\theta_w} \le \mathbb{P}\left(\sup_{0\le t\le h} Q_\Lambda(t) \ge w\right) \le (h+2)\theta_w.$$

Combining this with the fixed-interval expansion yields

$$\frac{h}{h+2} \le \liminf_{w\to\infty} \frac{\theta_w}{r(w)} \le \limsup_{w\to\infty} \frac{\theta_w}{r(w)} \le 1.$$

Letting $h \to \infty$ then shows that $\theta_w \sim r(w)$. Choose

$$w_n(u;\Lambda) = \lambda_1\{2\log\log n + \log\log\log n + 2u\}, \qquad \kappa_\Lambda = \sqrt{\frac{\lambda_1}{\pi(\lambda_1 - \lambda_0)}}.$$

Then $(\log n) r(w_n(u;\Lambda)) \to \kappa_\Lambda e^{-u}$. Taking $T = c\log n$ in the survival bounds now proves the stated limit for strict sublevels. Interchanging the coordinates covers $\lambda_0 > \lambda_1$.

For either choice of weights,

$$w_n(u+\eta;\Lambda) - w_n(u;\Lambda) = 2\lambda_\star \eta.$$

Monotonicity in u and continuity of $\exp(-c\kappa_\Lambda e^{-u})$ also make convergence uniform on compact u-sets. Squeezing at $u \pm \eta$ and then letting $\eta \downarrow 0$ covers non-strict sublevels and deterministic $o(1)$ threshold changes. Stationarity and the same squeeze at $c \pm \eta$ cover deterministic translates of intervals of length $(c + o(1))\log n$.

□

SA-5.2. Proof of Theorem SA-1. Write $\sigma^2 = \mathbb{V}[\varepsilon_i] > 0$, $V_i = \varepsilon_i/\sigma$, and $S_j(k) = \sum_{i=1}^k V_{\pi_j(i)}$. Let $\mathscr{I}(k,j)$ denote the root impurity gain at rank k in coordinate j. Then

$$\frac{\mathscr{I}(k,j)}{\sigma^2} = \left\{\frac{S_j(k) - (k/n)S_j(n)}{\sqrt{k(1-k/n)}}\right\}^2, \qquad 1 \le k < n.$$

The coordinate orderings are independent uniform permutations, independent of the errors, because the covariate coordinates are mutually independent, continuously distributed, and independent of the errors. Thus the criterion has the form required by Lemma SA-39, with $Q(v) = v^2$ and $J_n = \mathscr{I}/\sigma^2$.

One-coordinate approximation. The ordered V_i are i.i.d., centered, and standardized, with $\mathbb{E}[e^{h|V_1|}] < \infty$ for some $h > 0$. Dominated convergence then gives

$$\lambda\mathbb{E}[|V_1|^3 e^{\lambda|V_1|}] \to 0 < \mathbb{V}[V_1] = 1 \qquad (\lambda \downarrow 0),$$

so the Sakhanenko parameter is bounded below by a fixed positive constant. The scalar Komlós–Major–Tusnády (KMT) approximation [2, Theorem 2.6.2, p. 107], in the finite-sample form supplied by Sakhanenko's exponential inequality [23, Lemma 4.2], gives standard Gaussian partial sums $W_n(k)$ and fixed $c_{\mathrm{KMT}}, \lambda > 0$ such that

$$\mathbb{E}\left[\exp\left\{c_{\mathrm{KMT}}\lambda \max_{1\le k\le n} |S_1(k) - W_n(k)|\right\}\right] \le 1 + \lambda\sqrt{n},$$
$$\mathbb{P}\left(\max_{1\le k\le n} |S_1(k) - W_n(k)| > C_{\mathrm{KMT}} \log n\right) \le (1 + \lambda\sqrt{n}) n^{-c_{\mathrm{KMT}}\lambda C_{\mathrm{KMT}}} \to 0, \qquad C_{\mathrm{KMT}} > \frac{1}{2c_{\mathrm{KMT}}\lambda}.$$

Thus $\max_{1\le k\le n} |S_1(k) - W_n(k)| = O_{\mathbb{P}}(\log n)$. Let $\zeta_0, \dots, \zeta_{n-1}$ be independent standard Brownian bridges, independent of the Gaussian partial sums, and set

$$W_n(k + t) = (1 - t)W_n(k) + tW_n(k + 1) + \zeta_k(t), \qquad 0 \le k < n, \quad 0 \le t \le 1.$$

This is a continuous centered Gaussian process with covariance $\mathbb{E}[W_n(s)W_n(t)] = \min\{s, t\}$, hence a standard Wiener process [21, Section 3, Eq. (10), and Section 3.4]. Its values at integer times, and the approximation bound, are unchanged. For $B_n(t) = n^{-1/2}\{W_n(nt) - tW_n(n)\}$, this implies

$$\max_{r_n\le k\le n-r_n} \left| \frac{S_1(k) - (k/n)S_1(n)}{\sqrt{k(1 - k/n)}} - \frac{B_n(k/n)}{\sqrt{(k/n)(1 - k/n)}} \right| = O_{\mathbb{P}}\left(\frac{\log n}{\sqrt{r_n}}\right).$$

Let D_n denote the maximum absolute difference in the preceding display, and define

$$G_n = \max_{r_n\le k\le n-r_n} \frac{|B_n(k/n)|}{\sqrt{(k/n)(1 - k/n)}}.$$

Gaussian tails and the bound for D_n give

$$\mathbb{P}(G_n > x) \le 2ne^{-x^2/2}, \qquad G_n = O_{\mathbb{P}}(\sqrt{\log n}), \qquad D_n = O_{\mathbb{P}}(\log n/\sqrt{r_n}).$$

Since $|x^2 - y^2| \le |x - y|(2|y| + |x - y|)$, the squared-score error is at most

$$D_n(2G_n + D_n) = O_{\mathbb{P}}\left(\frac{(\log n)^{3/2}}{\sqrt{r_n}} + \frac{(\log n)^2}{r_n}\right) = o_{\mathbb{P}}(1).$$

This proves (SA-21).

For the endpoints, use the Darling–Erdős theorem as stated in [1, Theorem 2.1, pp. 121–122]. If X_i are i.i.d., centered, with $\mathbb{E}[X_i^2] = \sigma_X^2 > 0$ and $\mathbb{E}[|X_i|^3] < \infty$, put

$$M_r = \max_{1\le k\le r} \frac{|\sum_{i=1}^k X_i|}{\sigma_X\sqrt{k}}.$$

The absolute-maximum limit in that theorem implies

$$\sqrt{2\log\log r}\, M_r = 2\log\log r + \tfrac{1}{2}\log\log\log r + O_{\mathbb{P}}(1), \qquad \frac{M_r}{\sqrt{2\log\log r}} = 1 + O_{\mathbb{P}}\left(\frac{1 + \log\log\log r}{\log\log r}\right).$$

Therefore

$$\max_{1\le k\le r} \frac{(\sum_{i=1}^k X_i)^2}{k} = 2\sigma_X^2 \log\log r\{1 + o_{\mathbb{P}}(1)\}, \qquad r \to \infty. \tag{SA-23}$$

For $\sigma_X^2 = 0$ the maximum is zero almost surely. The subexponential assumption supplies the required third moment, so for every fixed coordinate j,

$$\max_{1\le k\le r_n} \frac{S_j(k)^2}{k} = O_{\mathbb{P}}(\log\log r_n) = O_{\mathbb{P}}(\log\log\log n).$$

For $k \le r_n < n/2$, the elementary bound

$$\frac{\{S_j(k) - (k/n)S_j(n)\}^2}{k(1 - k/n)} \le 4\frac{S_j(k)^2}{k} + 4\frac{r_n}{n}\frac{S_j(n)^2}{n}$$

controls the bridge correction. The last term is $O_{\mathbb{P}}(r_n/n)$ because $S_j(n)/\sqrt{n} = O_{\mathbb{P}}(1)$. Reversal gives the same bound at the other endpoint. A union bound over fixed p therefore gives

$$\max_{1\le j\le p} \max_{\substack{1\le k<r_n \\ \text{or } n-r_n<k<n}} \frac{\mathscr{I}(k,j)}{\sigma^2} = O_{\mathbb{P}}(\log\log r_n + r_n/n) = o_{\mathbb{P}}(L_n). \tag{SA-24}$$

Set

$$\beta_n = 2L_n + \log L_n - \log\pi, \qquad w_n(u) = \beta_n + 2u. \tag{SA-25}$$

Lemma SA-35 with $d = 1$ implies

$$\mathbb{P}\left(\sup_{0\le t\le c\log n} U(t)^2 \le w_n(u) \right) \to \exp(-ce^{-u}),$$

since $a(\log n)\sqrt{w_n(u)} - b_1(\log n) = u + o(1)$. Together with (SA-24), this verifies all one-coordinate and endpoint hypotheses of Lemma SA-39, with $A = 2$ and $\kappa = 1$. This step uses no assertion about the selected split location.

Fixed dimension and the winning block. Fix $0 < a < b < 1$ and $\ell \in [p]$. Partition the left boundary in coordinate ℓ at n^a, n^b. Its three length coefficients, followed by those of the other boundaries, are

$$a, \quad b-a, \quad 1-b, \quad \underbrace{1,\dots,1}_{2p-1}, \qquad a + (b-a) + (1-b) + (2p-1) = 2p.$$

The scalar intensity c, also appearing in the corrected Eicker law, becomes $b - a$ for the target after logarithmic time change. Lemmas SA-36 and SA-39 now supply the joint independent limits needed to compare the blocks. The proof of Lemma SA-39 also shows that the middle range and discarded endpoint ranks attain the overall maximum with probability tending to zero. Hence

$$\mathbb{P}\big(n^a \le \hat{\imath} \le n^b,\ \hat{\jmath} = \ell\big) \to \frac{b-a}{2p}.$$

The reflected target $[n - \lfloor n^b \rfloor, n - \lceil n^a \rceil]$ has the same coefficient. Lemma SA-37 then gives both limits for every measurable tie rule, including discrete errors.

For each $\ell \in [p]$, the left and right target events are disjoint for all sufficiently large n. Adding their joint limits and letting $a \downarrow 0$ and $b \uparrow 1$ gives $\liminf_{n\to\infty} \mathbb{P}(\hat{\jmath} = \ell) \ge 1/p$. Since the coordinate probabilities sum to one, $\mathbb{P}(\hat{\jmath} = \ell) \to 1/p$. Dividing the joint limits by this probability proves the conditional window limits. Adding these limits, in turn, gives

$$\mathbb{P}\left(a \le \frac{\log\min\{\hat{\imath}, n-\hat{\imath}\}}{\log n} \le b \,\middle|\, \hat{\jmath} = \ell \right) \to b - a.$$

Letting the interior endpoints approach zero and one excludes endpoint mass and proves the conditional uniform limit. Summing the joint limits over coordinates gives the corresponding unconditional limits. The same argument gives the full-maximum limits used in the estimation-error proofs,

$$\mathbb{P}\left(\max_{\substack{1\le k<n \\ 1\le j\le p}} \frac{\mathscr{I}(k,j)}{\sigma^2} \le \beta_n + 2u \right) \to \exp(-2pe^{-u}), \qquad \frac{\max_{\substack{1\le k<n \\ 1\le j\le p}} \mathscr{I}(k,j)}{2\sigma^2 \log\log n} \to_{\mathbb{P}} 1. \tag{SA-26}$$

SA-5.3. Proof of Theorem SA-2. Let $m = \min\{\hat{\imath}, n-\hat{\imath}\}$ and write $\mathtt{t}_S, \mathtt{t}_{S^c}$ for the smaller and larger selected root children. Since the CART score is a squared difference of child means, the exact identity

$$\sqrt{m}\,|\hat{\mu}^{\mathtt{stump}}(\mathtt{t}_S) - \hat{\mu}^{\mathtt{stump}}(\mathtt{t}_{S^c})| = \sqrt{\frac{n}{n-m}\mathscr{I}(\hat{\imath},\hat{\jmath})} \tag{SA-27}$$

holds. Here $\hat{\mu}^{\mathtt{stump}}(\mathtt{t})$ denotes the constant fitted value on the child $\mathtt{t}$. By (SA-17),

$$\mathbb{E}\left[\max_{1\le\ell\le p} \max_{n/2\le k\le n} \frac{\left|\sum_{i=1}^k \varepsilon_{\pi_\ell(i)}\right|^2}{k^2} \right] \le \frac{4}{n^2}\sum_{\ell=1}^p \mathbb{E}\left[\max_{1\le k\le n} \left|\sum_{i=1}^k \varepsilon_{\pi_\ell(i)}\right|^2 \right] \le \frac{16p\sigma^2}{n}.$$

Markov's inequality, applied also to the reversed orderings, gives

$$|\hat{\mu}^{\mathtt{stump}}(\mathtt{t}_{S^c}) - \mu| = O_{\mathbb{P}}(n^{-1/2}).$$

Equation (SA-26) gives $\mathscr{I}(\hat{\imath},\hat{\jmath}) = (2\sigma^2 + o_{\mathbb{P}}(1))L_n$.

Theorem SA-1 gives $\log m/\log n \rightsquigarrow \mathsf{Uniform}(0,1)$, hence $m/n \to_{\mathbb{P}} 0$ and $m/L_n \to_{\mathbb{P}} \infty$. The supremum error is the larger of the two child errors, so (SA-27) and the preceding bounds give

$$\frac{\sqrt{m}}{\sigma\sqrt{2L_n}} \sup_{\mathbf{x}\in\mathcal{X}} |\hat{\mu}^{\mathtt{stump}}(\mathbf{x}) - \mu(\mathbf{x})| \to_{\mathbb{P}} 1.$$

This proves uniform consistency. For fixed $b \in (0,1)$ and $0 < \epsilon < \min\{b, 1-b\}$, outside an event of probability $o(1)$,

$$\{m \le n^{b-\epsilon}\} \subseteq \left\{ \sup_{\mathbf{x}\in\mathcal{X}} |\hat{\mu}^{\mathtt{stump}}(\mathbf{x}) - \mu(\mathbf{x})| \ge \sigma n^{-b/2}\sqrt{2L_n} \right\} \subseteq \{m \le n^{b+\epsilon}\}.$$

The outer probabilities converge to $b-\epsilon$ and $b+\epsilon$. Letting $\epsilon \downarrow 0$ then proves (SA-4).

Fix $b \in (0,1)$. On $\{m \le n^b\}$, (SA-27) implies

$$\begin{aligned}
\left| \sqrt{m}\,|\hat{\mu}^{\mathtt{stump}}(\mathtt{t}_S) - \mu| - \sqrt{\mathscr{I}(\hat{\imath},\hat{\jmath})} \right| &\le \sqrt{m}\,|\hat{\mu}^{\mathtt{stump}}(\mathtt{t}_{S^c}) - \mu| + \left(\sqrt{\frac{n}{n-m}} - 1 \right) \sqrt{\mathscr{I}(\hat{\imath},\hat{\jmath})} \\
&\le n^{b/2} |\hat{\mu}^{\mathtt{stump}}(\mathtt{t}_{S^c}) - \mu| + C n^{b-1} \sqrt{\mathscr{I}(\hat{\imath},\hat{\jmath})} \\
&= O_{\mathbb{P}}(n^{(b-1)/2}) + O_{\mathbb{P}}(n^{b-1}\sqrt{L_n}) = o_{\mathbb{P}}(1).
\end{aligned}$$

Thus the triangle inequality and (SA-27) give, for every $\epsilon > 0$,

$$\mathbb{P}\left(m \le n^b,\ \left| \sqrt{m}\,|\hat{\mu}^{\mathtt{stump}}(\mathtt{t}_S) - \mu| - \sqrt{\mathscr{I}(\hat{\imath},\hat{\jmath})} \right| > \epsilon \right) \to 0.$$

The restriction $m \le n^b$ is essential in this display. Together with (SA-26), it implies that

$$T_n = \mathbb{1}(m \le n^b) \max\left\{ 2 - \frac{m|\hat{\mu}^{\mathtt{stump}}(\mathtt{t}_S) - \mu|^2}{\sigma^2 L_n},\ 0 \right\} \to_{\mathbb{P}} 0.$$

Choose a deterministic sequence $\delta_n \downarrow 0$ such that $\mathbb{P}(T_n > \delta_n) \to 0$. For example, choose strictly increasing integers N_j such that $\sup_{n \ge N_j} \mathbb{P}(T_n > 1/j) \le 1/j$, and set $\delta_n = 1/j$ for $N_j \le n < N_{j+1}$. Define, using this sequence, the root event

$$\mathcal{G}_n(b) = \left\{ m \le n^b, \quad \sqrt{m}\,|\hat{\mu}^{\mathtt{stump}}(\mathtt{t}_S) - \mu| \ge \sigma\sqrt{(2-\delta_n)L_n} \right\}. \tag{SA-28}$$

Then $\mathbb{P}(\{m \le n^b\} \setminus \mathcal{G}_n(b)) \to 0$. For every fixed $0 < a < b$, Theorem SA-1 gives $\mathbb{P}(n^a \le m \le n^b) \to b - a$, and hence $\liminf_{n\to\infty} \mathbb{P}(\mathcal{G}_n(b)) \ge b - a$. The sequence δ_n depends on b and the fixed model, but not on a. Letting $a \downarrow 0$ therefore gives

$$\liminf_{n\to\infty} \mathbb{P}(\mathcal{G}_n(b)) \ge b. \tag{SA-29}$$

The remaining boundary assertion requires translating split ranks into covariate locations. For i.i.d. $U_1, \dots, U_n \sim \mathsf{Uniform}(0,1)$, let $U_{(1)} < \cdots < U_{(n)}$ and π denote their ordered values and rank permutation. The joint density and beta marginals are given in [16, Chapter I, Theorem 4.3, pp. 17–18]. For every permutation ϖ and Borel set $A \subseteq \{0 < u_1 < \cdots < u_n < 1\}$,

$$\mathbb{P}\big(\pi = \varpi,\ (U_{(1)}, \dots, U_{(n)}) \in A\big) = \int_A \mathrm{d}u_1 \cdots \mathrm{d}u_n = \frac{1}{n!}\mathbb{P}\big((U_{(1)}, \dots, U_{(n)}) \in A\big). \tag{SA-30}$$

Thus the rank permutation is uniform and independent of the ordered values. Integrating the joint density over the other coordinates gives

$$f_{U_{(k)}}(u) = \frac{n!}{(k-1)!(n-k)!} u^{k-1}(1-u)^{n-k}, \qquad 0 < u < 1, \qquad \mathbb{E}[U_{(k)}^h] = \frac{k(k+1)\cdots(k+h-1)}{(n+1)(n+2)\cdots(n+h)}, \qquad h = 1, 2. \tag{SA-31}$$

Coordinate independence also makes (SA-30) hold jointly for the p coordinate orderings.

For the boundary assertion, fix $0 < a < b$ and a coordinate ℓ, and define

$$E_{\ell,n}^{\mathtt{L}} = \{n^a \le \hat{\imath} \le n^b,\ \hat{\jmath} = \ell\} \qquad \text{and} \qquad E_{\ell,n}^{\mathtt{R}} = \{n^a \le n - \hat{\imath} \le n^b,\ \hat{\jmath} = \ell\}.$$

Each event has probability $(b-a)/(2p) + o(1)$. Put $k_n = \lceil n^a \rceil$, $t_n = \eta_n n^{a-1}$, and $U_n = x_{\pi_\ell(k_n),\ell}$. By (SA-31),

$$\mathbb{E}[U_n] = \frac{k_n}{n+1}, \qquad \frac{\mathbb{V}[U_n]}{(\mathbb{E}[U_n])^2} = \frac{n - k_n + 1}{k_n(n+2)} = O(n^{-a}), \qquad \frac{t_n}{\mathbb{E}[U_n]} \to 0.$$

Chebyshev's inequality and reflection therefore give

$$\begin{aligned}
F_{\ell,n}^{\mathtt{L}} &= \{x_{\pi_\ell(k_n),\ell} \ge t_n\}, & \mathbb{P}((F_{\ell,n}^{\mathtt{L}})^c) &\le \frac{\mathbb{V}[U_n]}{(\mathbb{E}[U_n] - t_n)^2} = o(1), \\
F_{\ell,n}^{\mathtt{R}} &= \{1 - x_{\pi_\ell(n-k_n+1),\ell} > t_n\}, & \mathbb{P}((F_{\ell,n}^{\mathtt{R}})^c) &= o(1).
\end{aligned}$$

On $E^{\mathtt{L}}_{\ell,n} \cap F^{\mathtt{L}}_{\ell,n}$, the selected smaller child contains the entire strip $\{\mathbf{x} : x_\ell \le t_n\}$. On $E^{\mathtt{R}}_{\ell,n} \cap F^{\mathtt{R}}_{\ell,n}$, the selected cut rank is at most $n - k_n$, and its threshold is strictly below $1 - t_n$. Its smaller child therefore contains the entire strip $\{\mathbf{x} : 1 - x_\ell \le t_n\}$. Intersecting either event with $\mathcal{G}_n(b)$ gives the asserted error bound throughout its respective strip. The probabilities lost by either intersection tend to zero, without any independence assumption.

Every point of $\mathcal{X}_n(a, \eta_n)$ belongs to at least one boundary strip. With the same δ_n for all $2p$ strips,

$$\inf_{\mathbf{x} \in \mathcal{X}_n(a,\eta_n)} \mathbb{P}\left(|\hat{\mu}^{\mathtt{stump}}(\mathbf{x}) - \mu(\mathbf{x})| \ge \sigma n^{-b/2} \sqrt{(2-\delta_n) L_n} \right) \ge \min_{1 \le \ell \le p,\ s \in \{\mathtt{L},\mathtt{R}\}} \mathbb{P}(E^s_{\ell,n} \cap F^s_{\ell,n} \cap \mathcal{G}_n(b)) \to \frac{b-a}{2p}.$$

This proves (SA-5).

SA-5.4. Proof of Theorem SA-3. Let

$$\bar{\varepsilon}(\mathtt{t}) = \frac{1}{n(\mathtt{t})} \sum_{\mathbf{x}_i \in \mathtt{t}} \varepsilon_i$$

denote the centered average over a node. Use the deterministic sequence δ_n and event $\mathcal{G}_n(b)$ from (SA-28). Its limiting lower probability is at least b by (SA-29), and on this event the smaller root child $\mathtt{t} = \mathtt{t}_S$ satisfies

$$|\bar{\varepsilon}(\mathtt{t})| \ge \sigma n^{-b/2} \sqrt{(2-\delta_n) \log\log n}.$$

Any later recursive splitting only refines this root child. If $\mathcal{P}(\mathtt{t})$ denotes the collection of terminal descendants of $\mathtt{t}$ in the final tree, with $\mathcal{P}(\mathtt{t}) = \{\mathtt{t}\}$ if the child is never split again, then

$$\bar{\varepsilon}(\mathtt{t}) = \sum_{\mathtt{t}' \in \mathcal{P}(\mathtt{t})} \frac{n(\mathtt{t}')}{n(\mathtt{t})} \bar{\varepsilon}(\mathtt{t}').$$

Consequently, on $\mathcal{G}_n(b)$,

$$\sup_{\mathbf{x} \in \mathcal{X}} |\hat{\mu}^{\mathtt{NSS}}(\mathbf{x}) - \mu(\mathbf{x})| \ge \max_{\mathtt{t}' \in \mathcal{P}(\mathtt{t})} |\bar{\varepsilon}(\mathtt{t}')| \ge |\bar{\varepsilon}(\mathtt{t})| \ge \sigma n^{-b/2} \sqrt{(2-\delta_n) \log\log n}.$$

Equation (SA-29) then proves the theorem.

SA-5.5. Proof of Theorem SA-4. Work on the normalized cube $[0,1]^p$ with $n \ge 2$. As in the notation section, C may change from line to line and depends only on the error law, the fixed dimension p, and the fixed exponent q. We prove the bounds simultaneously for all axis-aligned rectangles $\mathtt{t}$, which allows the leaves to be chosen using the outcomes.

For the mass bound, let $\mathcal{Q}_n$ be the finite collection of open rectangles with endpoints in the grid $\{j/n^3 : 0 \le j \le n^3\}$, including the empty rectangle. There are then at most $(n^3+1)^{2p} + 1$ such rectangles. Every rectangle $\mathtt{t} \subseteq [0,1]^p$ contains some $Q \in \mathcal{Q}_n$, possibly empty, with

$$P_X(\mathtt{t}) - P_X(Q) \le 2p/n^3.$$

Round each lower endpoint up and each upper endpoint down to this grid, and use the open interval between the rounded endpoints. Each coordinate therefore loses length at most $2/n^3$. This inclusion holds regardless of which endpoints belong to $\mathtt{t}$. For a fixed grid rectangle, $n(Q)$ is binomial with mean $nP_X(Q)$, and the binomial lower-tail bound (SA-19) gives

$$\mathbb{P}\big(n(Q) < nP_X(Q)/2\big) \le \exp\{-nP_X(Q)/8\}.$$

A union bound over grid rectangles with $nP_X(Q) \ge L \log n$, for a sufficiently large fixed L depending only on p, gives

$$\mathcal{E}^{(1)}_n = \bigcap_{Q \in \mathcal{Q}_n} \{nP_X(Q) \le 2n(Q) + L \log n\}, \qquad \mathbb{P}((\mathcal{E}^{(1)}_n)^c) \le Cn^{-6}.$$

Indeed, a violation requires both $nP_X(Q) > L \log n$ and $n(Q) < nP_X(Q)/2$, so taking $L > 8(6p+6)$ suffices after enlarging C. Since $Q \subseteq \mathtt{t}$, this event implies

$$P_X(\mathtt{t}) \le C \frac{n(\mathtt{t}) + \log n}{n} \qquad \text{for every rectangle } \mathtt{t}.$$

Define, to control the residual sums, the finite family of observation subsets

$$\mathscr{I}_n = \big\{ \{i : \mathbf{x}_i \in \mathtt{t}\} : \mathtt{t} \subseteq [0,1]^p \text{ is an axis-aligned rectangle} \big\}.$$

Conditional on all sample covariates, this family is fixed and has at most $(n+1)^{2p}$ elements. Each coordinate interval selects a consecutive segment in its coordinate ordering, specified by two cuts among $n+1$ positions. The errors, under this conditioning,

remain independent with their original centered subexponential law. Thus (SA-18), or [20, Theorem 2.8.1], gives, for every $I \in \mathscr{I}_n$ of size $m \geq 1$ and every $t > 0$,

$$\mathbb{P}\left(\left| \sum_{i \in I} \varepsilon_i \right| > C\{\sqrt{mt} + t\} \,\middle|\, \mathbf{x}_1, \dots, \mathbf{x}_n \right) \leq 2e^{-t}.$$

Take $t = (2p+7)\log n$ and enlarge a fixed constant C_0 to absorb this multiple in the threshold. The conditional union bound, followed by averaging over the covariates, gives

$$\mathcal{E}_n^{(2)} = \bigcap_{\substack{I \in \mathscr{I}_n \\ |I|>0}} \left\{ \left| \sum_{i \in I} \varepsilon_i \right| \leq C_0\big(\sqrt{|I| \log n} + \log n\big) \right\}, \qquad \mathbb{P}((\mathcal{E}_n^{(2)})^c) \leq Cn^{-6}.$$

Both events are measurable finite intersections, since the sample orderings determine $\mathscr{I}_n$. Their intersection $\mathcal{E}_n = \mathcal{E}_n^{(1)} \cap \mathcal{E}_n^{(2)}$ satisfies $\mathbb{P}(\mathcal{E}_n^c) \leq Cn^{-6}$. On $\mathcal{E}_n$, the mass bound and

$$\left| \frac{1}{n(\mathsf{t})} \sum_{i : \mathbf{x}_i \in \mathsf{t}} \varepsilon_i \right|^2 \leq C \frac{\log^2 n}{n(\mathsf{t})} \qquad \text{whenever } n(\mathsf{t}) > 0$$

hold simultaneously for all rectangles.

First suppose $q \geq 2$. Let $\mathcal{P}$ be the terminal partition of any tree in the statement. Each leaf is nonempty in the sample. On $\mathcal{E}_n$, its contribution to integrated error satisfies

$$P_X(\mathsf{t}) \left| \frac{1}{n(\mathsf{t})} \sum_{i : \mathbf{x}_i \in \mathsf{t}} \varepsilon_i \right|^q \leq \frac{C \log^q n}{n} \left\{ n(\mathsf{t})^{1-q/2} + \log n \, n(\mathsf{t})^{-q/2} \right\} \leq C \frac{\log^{q+1} n}{n}.$$

Summing over at most 2^K leaves therefore bounds the integrated q-th-power error by $C 2^K \log^{q+1} n / n$ on $\mathcal{E}_n$. On every sample, that error is at most $\max_{1 \leq i \leq n} |\varepsilon_i|^q$. The exponential moment assumption implies $\mathbb{E}[|\varepsilon_i|^{2q}] < \infty$, so Cauchy–Schwarz gives

$$\mathbb{E}\left[\int_{\mathcal{X}} |\hat{\mu}^{\mathtt{NSS}}(\mathbf{x}) - \mu(\mathbf{x})|^q \, \mathrm{d}F_X(\mathbf{x}) \, \mathbb{1}(\mathcal{E}_n^c) \right] \leq \left\{ n \mathbb{E}[|\varepsilon_i|^{2q}] \mathbb{P}(\mathcal{E}_n^c) \right\}^{1/2} \leq Cn^{-5/2}.$$

Thus, for $q \geq 2$,

$$\mathbb{E}[\|\hat{\mu}^{\mathtt{NSS}} - \mu\|_q^q] \leq C \frac{2^K \log^{q+1} n}{n} + Cn^{-5/2} \leq C \frac{2^K \log^{q+1} n}{n}.$$

For $1 \leq q < 2$, Hölder's and Jensen's inequalities give

$$\|\hat{\mu}^{\mathtt{NSS}} - \mu\|_q \leq \|\hat{\mu}^{\mathtt{NSS}} - \mu\|_2, \qquad \mathbb{E}[\|\hat{\mu}^{\mathtt{NSS}} - \mu\|_q^q] \leq \left(\mathbb{E}[\|\hat{\mu}^{\mathtt{NSS}} - \mu\|_2^2] \right)^{q/2} \leq C \left(\frac{2^K \log^3 n}{n} \right)^{q/2}.$$

Taking q-th roots then proves the moment bounds. For either range of q, the corresponding inequalities on $\mathcal{E}_n$ give, for a sufficiently large fixed C_2,

$$\left\{ \|\hat{\mu}^{\mathtt{NSS}} - \mu\|_q \geq C_2 \left(\frac{2^K \log^{\max\{3,q+1\}} n}{n} \right)^{1/\max\{2,q\}} \right\} \subseteq \mathcal{E}_n^c, \qquad \mathbb{P}(\mathcal{E}_n^c) \leq Cn^{-6}.$$

The rectangle bounds hold before the partition is selected, for every deterministic depth bound $K = K_n$.

SA-5.6. Proof of Theorem SA-5. We first record a finite-count bound that also applies to honest causal averages.

Lemma SA-42 (Finite-count anti-concentration)**.** *Let $X_1, \dots, X_m$ be independent centered random variables with finite fourth moments satisfying $\mathbb{E}[X_i^4] \leq \kappa \mathbb{E}[X_i^2]^2$ for some $\kappa \geq 3$. For deterministic real weights a_i, put $S = \sum_{i=1}^m a_i X_i$ and $v = \mathbb{E}[S^2] \in (0, \infty)$. Then*

$$\mathbb{E}[S^4] \leq \kappa v^2, \qquad \mathbb{P}(|S| \geq \sqrt{v/2}) \geq \frac{1}{4\kappa}.$$

Proof. Independence and centering give

$$\mathbb{E}[S^4] = 3v^2 + \sum_{i=1}^m a_i^4 \{\mathbb{E}[X_i^4] - 3\mathbb{E}[X_i^2]^2\} \leq 3v^2 + (\kappa - 3)v^2.$$

Cauchy–Schwarz, in turn, gives the required Paley–Zygmund bound,

$$v = \mathbb{E}[S^2] \leq \frac{v}{2} + \mathbb{E}[S^2 \mathbb{1}(S^2 \geq v/2)] \leq \frac{v}{2} + \{\mathbb{E}[S^4] \mathbb{P}(S^2 \geq v/2)\}^{1/2}.$$

Rearranging then gives $\mathbb{P}(S^2 \geq v/2) \geq v^2/(4\mathbb{E}[S^4]) \geq 1/(4\kappa)$, as claimed. □

Throughout the proof, abbreviate the honest estimator $\hat{\mu}^{\texttt{HON}}(\mathbf{x})$ by $\check{\mu}(\mathbf{x})$. Let $M = n_{\mathsf{T}}$ and $N = n_\mu$ be the construction and estimation sample sizes, so $M + N = n$. Write $\tilde{\mathbf{x}}_1, \ldots, \tilde{\mathbf{x}}_N$ for the estimation covariates and $\tilde{\mathbf{X}}$ for their matrix. The fold-size assumption gives $\rho \le M/N \le \rho^{-1}$, and in particular both folds tend to infinity. Let $(\hat{\imath}, \hat{\jmath})$ be the root splitting index and coordinate selected from the construction sample, and let

$$\mathtt{t}_{\mathsf{L}} = \{\mathbf{x} \in \mathcal{X} : x_{\hat{\jmath}} \le x_{\pi_{\hat{\jmath}}(\hat{\imath}),\hat{\jmath}}\}$$

be the corresponding left root child. Define the number of estimation fold observations falling in the root child by

$$\tilde{\imath} = \sum_{i=1}^{N} \mathbb{1}(\tilde{x}_{i,\hat{\jmath}} \le x_{\pi_{\hat{\jmath}}(\hat{\imath}),\hat{\jmath}}).$$

To bound $\tilde{\imath}$, condition on $(\hat{\jmath}, \hat{\imath}) = (j, k)$. Let F_j denote the marginal distribution function of $x_{i,j}$ and put $U_{(k),j} = F_j(x_{\pi_j(k),j})$. Equation (SA-30) and coordinate independence in Assumption SA-1 make the rank permutations jointly independent of the coordinate order-statistic vectors. The errors are also independent of both by the same assumption. The observed-value threshold convention and the prescribed coordinate/cut-rank tie rule make the selected root coordinate and rank functions only of the permutations and errors. Thus, given $(\hat{\jmath}, \hat{\imath}) = (j, k)$ on any such event of positive probability, $U_{(k),j}$ has its original $\mathsf{Beta}(k, M - k + 1)$ law. Given $U_{(k),j}$, the count $\tilde{\imath}$ is $\mathsf{Binomial}(N, U_{(k),j})$, so its mean and variance are

$$\mathbb{E}\big[\,\tilde{\imath} \mid \hat{\jmath} = j,\, \hat{\imath} = k\,\big] = \frac{Nk}{M+1}, \qquad \mathbb{V}\big[\,\tilde{\imath} \mid \hat{\jmath} = j,\, \hat{\imath} = k\,\big] = \frac{Nk(M-k+1)(M+N+1)}{(M+1)^2(M+2)}. \tag{SA-32}$$

The variance follows from

$$\mathbb{E}[U_{(k),j}(1 - U_{(k),j})] = \frac{k(M-k+1)}{(M+1)(M+2)}, \qquad \mathbb{V}[U_{(k),j}] = \frac{k(M-k+1)}{(M+1)^2(M+2)},$$
$$\mathbb{V}[\tilde{\imath} \mid \hat{\jmath} = j,\, \hat{\imath} = k] = N\mathbb{E}[U_{(k),j}(1 - U_{(k),j})] + N^2\mathbb{V}[U_{(k),j}].$$

Writing $\mathrm{B}(\alpha, \beta) = \Gamma(\alpha)\Gamma(\beta)/\Gamma(\alpha + \beta)$ for the beta function gives

$$\mathbb{P}(\tilde{\imath} = 0 \mid \hat{\jmath} = j,\, \hat{\imath} = k) = \frac{\mathrm{B}(k, M - k + N + 1)}{\mathrm{B}(k, M - k + 1)} = \prod_{h=0}^{k-1} \frac{M-h}{M+N-h} \le \frac{M}{M+N}.$$

Hence $\mathbb{P}(\tilde{\imath} \ge 1 \mid \hat{\jmath} = j, \hat{\imath} = k) \ge N/(M+N)$ for every $k \ge 1$. Also, for $k \le M^b$, Markov's inequality and the conditional mean in (SA-32) give $\mathbb{P}(\tilde{\imath} > CM^b \mid \hat{\jmath} = j, \hat{\imath} = k) \le N/\{C(M+1)\}$. Subtracting these bounds then proves

$$\mathbb{P}\big(1 \le \tilde{\imath} \le CM^b \mid \hat{\jmath} = j, \hat{\imath} = k\big) \ge \frac{N}{M+N} - \frac{N}{C(M+1)}.$$

For example, take

$$C = \frac{2(1+\rho)}{\rho^2}, \qquad c_0 = \frac{\rho}{2(1+\rho)}.$$

Since $N/(M+N) \ge \rho/(1+\rho)$ and $N/(M+1) \le \rho^{-1}$, the last conditional probability is at least c_0, uniformly over j and $1 \le k \le M^b$. Define

$$\mathcal{E}_n = \{\hat{\imath} \le M^b,\ 1 \le \tilde{\imath} \le CM^b\}.$$

Theorem SA-1, summed over coordinates, gives $\liminf_{n\to\infty} \mathbb{P}(\hat{\imath} \le M^b) \ge (b-a)/2$ for every $0 < a < b$. Letting $a \downarrow 0$ and summing the preceding conditional bound over the selected ranks and coordinates yield

$$\liminf_{n\to\infty} \mathbb{P}(\mathcal{E}_n) \ge c_0 b/2.$$

Let $\mathscr{G} = \sigma(\mathcal{D}_{\mathsf{T}}, \tilde{\mathbf{X}})$. On $\mathcal{E}_n$, let $i_\star = \min\{1 \le i \le N : \tilde{\mathbf{x}}_i \in \mathtt{t}_{\mathsf{L}}\}$, let $\mathtt{t}_\star$ be its terminal cell, and put $\mathbf{x}_\star = \tilde{\mathbf{x}}_{i_\star}$ and $\tilde{\imath}_\star = \sum_{i=1}^{N} \mathbb{1}(\tilde{\mathbf{x}}_i \in \mathtt{t}_\star)$. On $\mathcal{E}_n^c$, define $i_\star = 1$, $\mathbf{x}_\star = \mathbf{0}$, let $\mathtt{t}_\star$ be the terminal cell containing $\mathbf{0}$, and set $\tilde{\imath}_\star = 1$ by convention. All four variables are then $\mathscr{G}$-measurable on the whole sample space. In the count and anti-concentration arguments below, restrict attention to $\mathcal{E}_n$, where the selected cell descends from the left root child and $1 \le \tilde{\imath}_\star \le \tilde{\imath} \le CM^b \le Cn^b$. On this event, given $\mathscr{G}$, the estimation residuals in this cell remain independent with their original error law. Set $\sigma_\mu^2 = \mathbb{V}[y_i] > 0$, $\kappa = \max\{3, \mathbb{E}[\varepsilon_i^4]/\sigma_\mu^4\}$, and $c_\mu = 1/(4\kappa)$. Applying Lemma SA-42 to their average gives, almost surely on $\mathcal{E}_n$,

$$\mathbb{P}\left(|\check{\mu}(\mathbf{x}_\star) - \mu(\mathbf{x}_\star)| \ge \frac{\sigma_\mu}{\sqrt{2\tilde{\imath}_\star}} \,\middle|\, \mathscr{G}\right) \ge c_\mu, \qquad \mathbf{x}_\star \in \mathtt{t}_\star.$$

Thus, since $\sigma_\mu = \sigma$, $\tilde{\imath}_\star \le Cn^b$ on $\mathcal{E}_n$, and $(2C)^{-1/2} = \rho/(2\sqrt{1+\rho})$,

$$\mathbb{P}\left(\sup_{\mathbf{x}\in\mathcal{X}} |\check{\mu}(\mathbf{x}) - \mu(\mathbf{x})| \ge \frac{\rho\sigma_\mu}{2\sqrt{1+\rho}} n^{-b/2}\right) \ge \mathbb{E}\left[\mathbb{1}(\mathcal{E}_n)\mathbb{P}\left(\,|\check{\mu}(\mathbf{x}_\star) - \mu(\mathbf{x}_\star)| \ge \sigma_\mu/\sqrt{2\tilde{\imath}_\star} \,\middle|\, \mathscr{G}\,\right)\right] \ge c_\mu \mathbb{P}(\mathcal{E}_n).$$

Taking a limit inferior finally gives a lower bound $c_\mu c_0 b/2 = \rho\sigma_\mu^4 b/\{16(1+\rho)\max\{3\sigma_\mu^4, \mathbb{E}[\varepsilon_i^4]\}\}$.

SA-5.7. Proof of Theorem SA-6. We first prove the bound for $q = 2$. Condition on the construction fold $\mathcal{D}_{\mathsf{T}}$, so its terminal partition $\mathcal{P}$ is fixed and independent of the estimation fold. Write $m = n_\mu$. For a leaf $\mathtt{t}$ of mass $u = P_X(\mathtt{t}) > 0$, its estimation count $B = n(\mathtt{t})$ has the conditional distribution $\mathsf{Binomial}(m, u)$. Given the estimation covariates and $B > 0$, the leaf average has mean μ and variance σ^2/B. When $B = 0$, the estimate is zero. Hence, for $\mathbf{x} \in \mathtt{t}$,

$$u\mathbb{E}\big[(\hat{\mu}^{\texttt{HON}}(\mathbf{x}) - \mu(\mathbf{x}))^2 \mid \mathcal{D}_{\mathsf{T}}\big] = u\sigma^2 \sum_{k=1}^{m} \frac{\mathbb{P}(B = k \mid \mathcal{D}_{\mathsf{T}})}{k} + u\mu^2(1-u)^m.$$

Integration gives the binomial reciprocal-moment identity

$$\mathbb{E}\left[\frac{1}{B+1}\,\middle|\,\mathcal{D}_{\mathsf{T}}\right] = \int_0^1 (1-u+ut)^m \,\mathrm{d}t = \frac{1-(1-u)^{m+1}}{(m+1)u}.$$

Since $1/k \le 2/(k+1)$ for $k \ge 1$, and $\mathbb{1}(B=0) \le 1/(B+1)$, this gives

$$u\sum_{k=1}^{m} \frac{\mathbb{P}(B = k \mid \mathcal{D}_{\mathsf{T}})}{k} \le 2u\mathbb{E}\left[\frac{1}{B+1}\,\middle|\,\mathcal{D}_{\mathsf{T}}\right] \le \frac{2}{m+1}, \qquad u(1-u)^m \le u\mathbb{E}\left[\frac{1}{B+1}\,\middle|\,\mathcal{D}_{\mathsf{T}}\right] \le \frac{1}{m+1}.$$

Leaves with $u = 0$ contribute zero to integrated risk. Summing over the partition therefore yields

$$\mathbb{E}\left[\int_{\mathcal{X}} (\hat{\mu}^{\texttt{HON}}(\mathbf{x}) - \mu(\mathbf{x}))^2 \,\mathrm{d}F_X(\mathbf{x})\,\middle|\,\mathcal{D}_{\mathsf{T}}\right] \le \frac{2\sigma^2+\mu^2}{m+1}|\mathcal{P}| \le \frac{2\sigma^2+\mu^2}{m+1}2^K.$$

The deterministic depth bound then permits averaging over the construction fold. Taking the square root proves the moment bound for $q = 2$ with $C_1 = (2\sigma^2+\mu^2)^{1/2}$. Applying Markov's inequality to integrated squared error and taking square roots also gives the probability bound with $C_2 = C_1$.

For fixed $q > 2$, Bernstein's inequality for the centered responses gives, for $k \ge 1$ and $t > 0$,

$$\mathbb{P}\left(\left|\frac{1}{k}\sum_{i=1}^{k}\varepsilon_i\right| > t\right) \le 2\exp\{-k\min\{t^2, t\}/C\}.$$

This follows from (SA-18) and is also stated in [20, Corollary 2.8.3]. Integrating this tail bound then yields

$$\mathbb{E}\left[\left|\frac{1}{k}\sum_{i=1}^{k}\varepsilon_i\right|^q\right] \le 2q\int_0^\infty t^{q-1}\{e^{-kt^2/C} + e^{-kt/C}\}\,\mathrm{d}t \le C(k^{-q/2} + k^{-q}) \le Ck^{-q/2},$$

where C may now depend on q. Since $k^{-q/2} \le 1/k$, the same binomial bounds give, for $\mathbf{x} \in \mathtt{t}$,

$$u\mathbb{E}\big[|\hat{\mu}^{\texttt{HON}}(\mathbf{x}) - \mu(\mathbf{x})|^q \mid \mathcal{D}_{\mathsf{T}}\big] \le Cu\sum_{k=1}^{m}\frac{\mathbb{P}(B=k\mid\mathcal{D}_{\mathsf{T}})}{k} + u|\mu|^q\mathbb{P}(B=0\mid\mathcal{D}_{\mathsf{T}}) \le \frac{C}{m+1}.$$

Summing over at most 2^K leaves, unconditioning, and taking the q-th root proves the moment bound. Markov's inequality applied to integrated q-th-power error, followed by taking q-th roots of the threshold, similarly proves the probability bound.

For $1 \le q < 2$, Hölder's and Jensen's inequalities give

$$\|\hat{\mu}^{\texttt{HON}} - \mu\|_q \le \|\hat{\mu}^{\texttt{HON}} - \mu\|_2, \qquad \big(\mathbb{E}[\|\hat{\mu}^{\texttt{HON}} - \mu\|_q^q]\big)^{1/q} \le \big(\mathbb{E}[\|\hat{\mu}^{\texttt{HON}} - \mu\|_2^2]\big)^{1/2} \le (2\sigma^2+\mu^2)^{1/2}\left(\frac{2^K}{m+1}\right)^{1/2}.$$

The event inclusion

$$\{\|\hat{\mu}^{\texttt{HON}} - \mu\|_q > t\} \subseteq \{\|\hat{\mu}^{\texttt{HON}} - \mu\|_2 > t\}$$

gives the same probability bound and constants as for $q = 2$. These calculations condition only on the independent construction fold and hold for every deterministic $K = K_n$.

SA-5.8. Proof of Theorem SA-7. Write $N = n_{\mathsf{T}}$, $M = n_\mu$, and $\sigma^2 = \mathbb{V}[\varepsilon_i] > 0$. Under Assumption SA-1, the residuals form a centered i.i.d. sequence, independent of the scalar covariate, with a moment generating function finite on a neighborhood of zero. Apply the probability integral transform described after Assumption SA-1, and order the construction observations by their transformed covariates. In this order set $Z_i = \varepsilon_i/\sigma$ and $S_j = \sum_{i=1}^{j} Z_i$, with $S_0 = 0$. The sequence $(Z_i)_{i=1}^N$ is thus i.i.d. and independent of the numerical covariate order statistics. For a node containing the consecutive ranks $a+1, \dots, a+m$, its standardized CART score at the jth cut is

$$\frac{|S_{a+j} - S_a - (j/m)(S_{a+m} - S_a)|}{\sqrt{j(1-j/m)}}, \qquad 1 \le j < m.$$

Maximizing this score is equivalent to minimizing the CART sum of squared errors.

The proof has three steps. Gaussian comparison first gives, with probability bounded away from zero,

$$\text{a node of size } O((\log N)^B) \quad \text{at depth } O(\log\log N)$$

for a fixed constant $B > 0$ (Lemma SA-47). Following smaller children then reaches a singleton in a further $O(\log\log N)$ splits. Its beta spacing law and the independent estimation fold finally give a fixed positive probability of error.

Gaussian histories contain only revealed Gaussian endpoints, winning cuts, and discarded boundary increments. They, in particular, do not include coupling success. The original residual law may be discrete or continuous.

Gaussian conditional bridges and positive excursions. Conditioning on Gaussian endpoints or winning heights is understood through regular conditional distributions [18, Theorems 2.1.22 and 4.1.18]. These exist because the paths and histories here take values in finite-dimensional Euclidean spaces, with finitely many discrete rank choices. All conditional assertions are accordingly for almost every realized history.

In the Gaussian arguments, put $L_m = \log\log(m \vee 3)$, $v_m(u) = 2L_m + 3\log L_m + 2u$, and $A_m(j) = \sqrt{j(1-j/m)}$. A positive Gaussian excursion E_m is the bridge of m independent standard normal increments from zero to zero, conditioned to be nonnegative at its internal integer ranks. The conditioning event has positive probability under the endpoint bridge law. Since nonnegativity is imposed only at integer ranks, this is a discrete excursion, distinct from the continuous Brownian excursion used below.

Lemma SA-43 (Scalar Gaussian histories)**.** *Consider independent standard normal increments in a fixed scalar ordering. After revealing their total sum, explore intervals sequentially, revealing each maximizing scalar cut rank and its signed winning height, with interval choices and stopping determined by the already revealed history. Boundary pieces may be discarded from an interval whose two endpoint sums are known, provided their lengths are chosen from the already revealed history and all their increments are revealed. The endpoint sums of every retained interval are then known as well. Conditional on this history, the unexplored disjoint intervals are independent Gaussian endpoint bridges subject to their inherited winning bands. Subtracting each interval's endpoint chord then gives density*

$$\frac{1}{c}\exp\left\{-\frac{1}{2}\sum_{j=1}^{m}(x_j - x_{j-1})^2\right\}\prod_{j=1}^{m-1}\mathbb{1}(\ell(j) \le x_j \le u(j)), \qquad x_0 = x_m = 0, \tag{SA-33}$$

where ℓ is convex, u is concave, and $\ell \le 0 \le u$. The normalizing integral c is positive for almost every realized history. The Gaussian kernel is, moreover, independent of the revealed mean when the inherited walls in (SA-33) are held fixed.

Proof. For an ancestor interval $[a, b]$, with chord q and winning score z, the winning inequalities are precisely

$$q(t) - z\sqrt{\frac{(t-a)(b-t)}{b-a}} \le T_t \le q(t) + z\sqrt{\frac{(t-a)(b-t)}{b-a}}, \qquad a < t < b,$$

where T denotes the Gaussian partial-sum path. For a cut k and signed standardized height h,

$$T_k = q(k) + h\sqrt{\frac{(k-a)(b-k)}{b-a}}, \qquad \left|\frac{\partial T_k}{\partial h}\right| = \sqrt{\frac{(k-a)(b-k)}{b-a}}.$$

Thus the Jacobian adds no path-dependent weight. Given the exposed sums and discarded increments, write p_j for the endpoint Gaussian density on the remaining interval I_j and $\mathcal{C}_j$ for its corridor. Each constraint involves one interval. Conditional on the revealed sigma-field $\mathcal{H}$,

$$p\left(\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(d)} \mid \mathcal{H}\right) = \prod_{j=1}^{d}\frac{p_j(\mathbf{x}^{(j)})\mathbb{1}(\mathbf{x}^{(j)} \in \mathcal{C}_j)}{\int_{\mathcal{C}_j} p_j(\mathbf{u})\,\mathrm{d}\mathbf{u}}.$$

Given the already revealed history, the interval choices and stopping decisions add no further conditioning, so the factorization applies at each step. Further boundary revelations preserve this factorization because the child endpoint sums are known. This applies to trimming at an exposed cut, but not to an interior block with unknown endpoint sums.

If q_I is the chord of a retained interval and ℓ_I, u_I are its inherited walls, the path satisfies the inherited bands at both endpoints, where it equals q_I; hence convexity and concavity give

$$\ell_I \le q_I \le u_I, \qquad \ell = \ell_I - q_I \le 0 \le u_I - q_I = u.$$

For an interval with mean μ_I, the exponent in its Gaussian density contains

$$\sum_{j=1}^{m} (\mu_I + x_j - x_{j-1})^2 = m\mu_I^2 + \sum_{j=1}^{m} (x_j - x_{j-1})^2.$$

Thus the Gaussian kernel in (SA-33) does not depend on μ_I, although the inherited walls may depend on the revealed endpoints. Under the density disintegration, the normalizing integral can vanish only on a null set. Distinct internal cut scores, finally, can tie only on a union of proper affine hyperplanes in the unexposed increments, which gives uniqueness of the Gaussian maximizing rank almost surely. □

Lemma SA-44 (Gaussian bridge ordering)**.** *For a Gaussian bridge subject to coordinatewise corridor constraints, increasing its endpoints, either corridor wall, or the coefficients of a linear field in its log density, all coordinatewise, increases the path in coordinatewise stochastic order whenever both laws are feasible. We write $X \le_{\mathrm{st}} Y$ when $\mathbb{E}[g(X)] \le \mathbb{E}[g(Y)]$ for every bounded coordinatewise nondecreasing measurable function g. In particular, a centered bridge X of the form (SA-33) satisfies $X \le_{\mathrm{st}} E_m$ and $-X \le_{\mathrm{st}} E_m$. These two comparisons, however, need not use the same coupling.*

Proof. Allow a term $\sum_{j=1}^{m-1} f_j x_j$ in the log density. Conditional on its neighboring heights a, b, the height at site j has normal density with mean $(a+b+f_j)/2$ and variance $1/2$, truncated to its allowed interval. Writing $\theta = (a+b+f_j)/2$, the untruncated densities satisfy

$$\frac{p_{\theta_2}(x)}{p_{\theta_1}(x)} = \exp\{2(\theta_2 - \theta_1)x + \theta_1^2 - \theta_2^2\}, \qquad \theta_1 \le \theta_2.$$

The ratio is therefore increasing in x. Restricting to a common interval preserves this property. If r is this normalized ratio and V, V' are independent with the first truncated law, then for every bounded increasing g,

$$\mathbb{E}[g(V)r(V)] - \mathbb{E}[g(V)] = \tfrac{1}{2}\mathbb{E}[(g(V) - g(V'))(r(V) - r(V'))] \ge 0.$$

Removing mass at the lower end or adding mass at the upper end also increases the conditional distribution. Thus increasing either neighbor, f_j, or either endpoint of the truncation interval increases the conditional distribution. On a bounded common mesh, update a uniformly selected site by the same uniform conditional-quantile variable in both chains. If $X_0^{(1)} \le X_0^{(2)}$, then

$$X_n^{(1)} \le X_n^{(2)} \quad (n \ge 0).$$

The conditional-update kernels preserve the discretized bridge laws π_1, π_2 [24, Section 3.3]. Single-site moves connect all allowed configurations, and positive holding probabilities give aperiodicity. Hence, with convergence in total variation,

$$\mathcal{L}(X_n^{(i)}) \to \pi_i \quad (i = 1, 2), \qquad \pi_1 \le_{\mathrm{st}} \pi_2$$

by [24, Theorem 4.9, p. 52]. Gaussian density integrals converge as the mesh decreases and the boxes increase. Their positive normalizers therefore preserve the ordering in the limit. Finally, increasing (ℓ, u) to $(0, \infty)$ gives $X \le_{\mathrm{st}} E_m$. Apply the same comparison to $-X$. □

We use positive-walk estimates in conjunction with these stochastic comparisons to bound split scores and localize nearly maximizing cuts. Let W_j be a random walk with independent standard normal increments, started at $x \ge 0$, and write G for one such increment. Put $\zeta = \inf\{j \ge 1 : W_j < 0\}$. Define

$$h(x) = x - \mathbb{E}_x[W_\zeta], \qquad x \ge 0.$$

Set $h(x) = 0$ for $x < 0$. The integer-time law of the iterated logarithm [2, Theorem 1.3.1 and Remark 1.3.1, p. 36] gives

$$\limsup_{j\to\infty} \frac{W_j}{\sqrt{2j \log\log j}} = 1, \qquad \liminf_{j\to\infty} \frac{W_j}{\sqrt{2j \log\log j}} = -1 \quad \text{almost surely.}$$

Hence $\zeta < \infty$ almost surely. For every fixed $r \ge 1$, the normal tail gives a uniform bound on the rth moment of an overshoot at either end of an interval. Indeed, writing $\overline{\Phi}$ for the standard normal survival function,

$$\frac{\overline{\Phi}(x+t)}{\overline{\Phi}(x)} = e^{-t^2/2}\mathbb{E}\left[e^{-tG} \,\middle|\, G > x\right] \le e^{-xt - t^2/2}, \qquad x, t \ge 0.$$

Conditioning on the last position before a crossing gives the same bound at a first crossing. Consequently, for an absolute $B_h < \infty$,

$$x \le h(x) \le x + B_h, \qquad h(x) \asymp 1 + x, \qquad h(x) = \mathbb{E}[h(x+G)\mathbb{1}(x+G \ge 0)]. \tag{SA-34}$$

For $0 \le x \le 1$,

$$h(x) \ge \mathbb{E}[-(x+G)\mathbb{1}(G < -2)] \ge \mathbb{P}(G < -2) > 0,$$

and $h(x) \ge x$ for $x \ge 1$. Conditioning at the first step and using $\mathbb{E}[G] = 0$ gives the harmonic identity.

For the survival and transition estimates, take $0 \le x \le R$ and put $\zeta_R = \inf\{j \ge 1 : W_j \notin [0, R]\}$. Apply optional stopping to the killed harmonic martingale and to $W_j^2 - j$, first at $\zeta_R \wedge n$ and then let $n \to \infty$ [18, Theorem 4.8.2, p. 256]. The moment and tail bounds below justify the passage to the limit and give

$$\begin{aligned} h(x) &= \mathbb{E}_x\left[h(W_{\zeta_R})\mathbb{1}(W_{\zeta_R} > R)\right] \ge R\mathbb{P}_x(W_{\zeta_R} > R), \\ \mathbb{E}_x[W_{\zeta_R}^2] &\le (R^2 + CR + C)\mathbb{P}_x(W_{\zeta_R} > R) + C \le CRh(x), \\ \mathbb{E}_x[\zeta_R] &= \mathbb{E}_x[W_{\zeta_R}^2] - x^2 \le CRh(x), \qquad R \ge 1. \end{aligned}$$

The second line follows from the first two overshoot bounds. For fixed R, an increment larger than R forces exit from every point of $[0, R]$. Independence and the fourth overshoot bound give

$$\mathbb{P}_x(\zeta_R > n) \le \{1 - \overline{\Phi}(R)\}^n, \qquad \sup_{n \ge 1} \mathbb{E}_x\left[|W_{\zeta_R \wedge n}|^4\right] \le C(1 + R^4).$$

Hence $\mathbb{E}_x[\zeta_R] < \infty$, the stopped squares are uniformly integrable, and $\zeta_R \wedge n \le \zeta_R$. These bounds justify both stopping limits. Taking $R = \sqrt{n}$, Markov's inequality and $\mathbb{E}_x[\zeta_R] \le CRh(x)$ give

$$\mathbb{P}_x(\zeta > n) \le \mathbb{P}_x(W_{\zeta_R} > R) + \mathbb{P}_x(\zeta_R > n) \le Ch(x)n^{-1/2}. \tag{SA-35}$$

For $x > \sqrt{n}$, the same inequality follows from $h(x) \ge x$.

Write $q_n(x, y)$ for the transition density of this walk killed on entering the negative half-line, and put

$$\phi_n(y) = (2\pi n)^{-1/2} \exp\{-y^2/(2n)\}.$$

Reversal gives $q_n(x, y) = q_n(y, x)$. For $n = n_1 + n_2 + n_3$ with $n_j \asymp n$, the Markov property and (SA-35) give

$$\begin{aligned} q_n(x, 0) &= \int_0^\infty \int_0^\infty q_{n_1}(x, u) q_{n_2}(u, v) q_{n_3}(v, 0)\, \mathrm{d}u\, \mathrm{d}v \\ &\le Cn^{-1/2}\mathbb{P}_x(\zeta > n_1)\mathbb{P}_0(\zeta > n_3) \le Ch(x)n^{-3/2}. \end{aligned}$$

Consequently,

$$q_n(x, 0) \le Ch(x)n^{-3/2}, \qquad q_n(0, 0) = \frac{\phi_n(0)}{n}. \tag{SA-36}$$

Increasing C also covers the short transitions. Conditional on a zero sum, the partial sums at ranks $0, \dots, n-1$ have a unique minimum. Exactly one cyclic rotation is nonnegative, and cyclic invariance gives

$$\mathbb{P}(\zeta > n - 1 \mid W_n = 0) = \frac{1}{n}, \qquad q_n(0, 0) = \phi_n(0)\mathbb{P}(\zeta > n - 1 \mid W_n = 0).$$

Let Q_x be the positive Doob transform with harmonic function h [24, Section 17.6.1, p. 256]. Its transition kernel is

$$Q(x, \mathrm{d}y) = \frac{h(y)}{h(x)}\phi_1(y - x)\mathbb{1}(y \ge 0)\, \mathrm{d}y, \qquad \int Q(x, \mathrm{d}y) = 1,$$

where the integral equals one by (SA-34). The prefix likelihood is then

$$\prod_{j=1}^{t} \frac{h(W_j)}{h(W_{j-1})}\mathbb{1}(W_j \ge 0) = \frac{h(W_t)}{h(x)}\mathbb{1}(\zeta > t).$$

By (SA-36), every nonnegative prefix functional Ψ of length $t \le m/2$ satisfies

$$\mathbb{E}[\Psi(E_m(0), \dots, E_m(t))] \le C\mathbb{E}_0^Q[\Psi(W_0, \dots, W_t)]. \tag{SA-37}$$

In fact the density ratio is $q_{m-t}(W_t, 0)h(0)/\{q_m(0, 0)h(W_t)\} \le C$.

Writing $h(x) = x + d(x)$, $0 \le d \le B_h$, gives

$$\mathbb{E}_x^Q\left[|W_1 - x|^r\right] \le C_r, \qquad |\mathbb{E}_x^Q[W_1 - x]| \le C/(1+x).$$

Indeed, the one-step density ratio is at most $C(1+|G|)$, and

$$h(x)\left|\mathbb{E}_x^Q[W_1 - x]\right| = |\mathbb{E}[G\{x + G + d(x+G)\}\mathbb{1}(G \ge -x)]| \le x\mathbb{E}[|G|\mathbb{1}(G < -x)] + \mathbb{E}[G^2] + B_h\mathbb{E}[|G|] \le C.$$

Divide by $h(x) \asymp 1 + x$. For integer $r \ge 1$, expansion then gives

$$\mathbb{E}_x^Q[W_1^{2r} - x^{2r}] \le C_r(1 + x^{2r-2}).$$

For $r \in \mathbb{N}$, put $\nu_y = \inf\{j \ge 0 : W_j > y\}$. Stopping and summing the preceding drift inequality gives

$$\mathbb{E}_0^Q\left[W_{t\wedge\nu_y}^{2r}\right] \le C_r \sum_{j=0}^{t-1}\{1 + \mathbb{E}_0^Q\left[W_{j\wedge\nu_y}^{2r-2}\right]\} \le C_r t^r,$$

where the last inequality follows by induction on r, starting with $r = 1$. Since $W_{t\wedge\nu_y} > y$ on $\{\nu_y \le t\}$, (SA-37) yields

$$\mathbb{P}\left(\max_{0\le j\le t} E_m(j) > y\right) \le C_r t^r y^{-2r}, \qquad 1 \le t \le m/2. \tag{SA-38}$$

Time reversal gives the corresponding bound at the other endpoint. Lemma SA-44 therefore implies, for every centered Gaussian corridor bridge and $H_0 > 0$,

$$\mathbb{P}\left(\max_{0\le j\le t}|X(j)| > H_0\sqrt{t}\,\middle|\,\mathcal{H}\right) \le C/H_0^2, \qquad t \le m/2.$$

All constants are, moreover, independent of the feasible scalar history $\mathcal{H}$.

We also use the return estimate

$$\mathbb{P}\left(\min_{R\le j\le\lfloor m/2\rfloor} E_m(j) \le y\right) \le C(1+y)/\sqrt{R}, \qquad 1 \le R \le m/2, \quad y \ge 0. \tag{SA-39}$$

To verify it, put $\nu_y^- = \inf\{j \ge 0 : W_j \in [0, y]\}$. For every integer $t \ge 0$, the stopped change of measure gives

$$Q_x(\nu_y^- \le t) = \frac{\mathbb{E}_x\left[h(W_{\nu_y^-})\mathbb{1}(\nu_y^- \le t,\ \nu_y^- < \zeta)\right]}{h(x)} \le \min\left\{1, \frac{C(1+y)}{h(x)}\right\}.$$

Letting $t \to \infty$ proves the same bound for $Q_x(\nu_y^- < \infty)$. Also,

$$\mathbb{E}_0^Q\left[\frac{1}{h(W_R)}\right] = \frac{\mathbb{P}_0(\zeta > R)}{h(0)} \le \frac{C}{\sqrt{R}}.$$

The Markov property at R then gives

$$Q_0\left(\min_{R\le j\le\lfloor m/2\rfloor} W_j \le y\right) \le C(1+y)\mathbb{E}_0^Q\left[\frac{1}{h(W_R)}\right] \le \frac{C(1+y)}{\sqrt{R}}.$$

Apply (SA-37), also in reversed time.

Lemma SA-45 (Gaussian excursion extremes)**.** *For every fixed $u \in \mathbb{R}$ and $a, \beta \in (0,1)$, as $m \to \infty$,*

$$\mathbb{P}\left(\max_{1\le j\le\lfloor m^a\rfloor}\frac{E_m(j)^2}{A_m(j)^2} > v_m(u)\right) \le Cae^{-u} + o(1), \tag{SA-40}$$

$$\mathbb{P}\left(\max_{m^\beta\le j\le m-m^\beta}\frac{E_m(j)^2}{A_m(j)^2} > v_m(u)\right) \le C(1-\beta)e^{-u} + o(1). \tag{SA-41}$$

The first bound also holds at the right endpoint. For every $A > \sqrt{2}$,

$$\mathbb{P}\left(\max_{1\le j<m}\frac{E_m(j)^2}{A_m(j)^2} > A^2 L_m\right) = o(1).$$

If $H = \lfloor m^\gamma\rfloor$, $0 < \vartheta < \gamma \le 1$, then

$$\mathbb{P}\left(\max_{\lceil m^\vartheta\rceil\le j\le H/(\log H)^4}\frac{E_H(j)^2}{j} \le v_m(u)\right) \le C\exp\{-c(\gamma-\vartheta)e^{-u}\} + o(1). \tag{SA-42}$$

The same bound holds with lower endpoint $(\log H)^{16}$ and $\gamma - \vartheta$ replaced by γ.

Proof. To compare the positive walk Q to a three-dimensional Bessel process without a strong approximation, realize the unconditioned Gaussian walk as the integer skeleton of Brownian motion $\mathcal{B}(t)$, and let $\overline{W}$ be its polygonal interpolation. For every fixed $D > 0$, the reflection bound (SA-20) and a union bound over the unit intervals give

$$\mathbb{P}\left(\sup_{0\le t\le n} \left|\overline{W}(t) - \mathcal{B}(t)\right| > b_n \right) \le C_D n^{-D}, \qquad b_n = C_D \log n.$$

Start both paths at $x > 0$. For a path event $\mathcal{E}$, write $\mathcal{E}^d$ for its enlargement in uniform distance by d. The terminal-weight definition of Q_x, together with $h(y) \le y + B_h$, implies

$$Q_x(\mathcal{E}) \le \frac{x + b_n + B_h}{h(x)} \mathbb{P}^{\mathrm{BES}(3)}_{x+b_n+B_h}\left(\mathcal{E}^{2b_n+B_h}\right) + \frac{C(x+\sqrt{n}+B_h)}{h(x)} n^{-D/2}. \tag{SA-43}$$

On interpolation success and discrete survival,

$$\inf_{0\le t\le n} \{\mathcal{B}(t) + b_n + B_h\} \ge 0, \qquad h(W_n) \le \mathcal{B}(n) + b_n + B_h.$$

The reflection principle [21, Section 6.1, pp. 38–39] and the Doob transform characterization of BES(3) with harmonic function $y \mapsto y$ [21, Section 7.1] give

$$k_t(x,y) = \phi_t(y-x) - \phi_t(y+x), \qquad p_t^{(3)}(x,y) = \frac{y}{x} k_t(x,y), \qquad x, y > 0.$$

This proves the first term. On interpolation failure, Cauchy–Schwarz gives the bound

$$\frac{\{\mathbb{E}_x\left[(|W_n| + B_h)^2\right]\}^{1/2}}{h(x)} (C_D n^{-D})^{1/2} \le \frac{C(x+\sqrt{n}+B_h)}{h(x)} n^{-D/2}.$$

Both calculations use the original joint law. Brownian motion, in particular, is not conditioned on interpolation success.

For a walk started at zero, set $x_n = (\log n)^2$ and $\nu_n = \inf\{j \ge 0 : W_j \ge x_n\}$. For the original Gaussian walk, a block of $\lceil 4x_n^2 \rceil$ increments has probability bounded below of displacement with absolute value greater than x_n, which forces exit from $[0, x_n]$. Independent blocks therefore give an exit-time tail Ce^{-ct/x_n^2}. On $\{\nu_n > t\}$ the terminal Doob density is at most $C(1 + x_n)$, so

$$Q_0(\nu_n > t) \le C(1 + x_n) e^{-ct/x_n^2}.$$

The increment density ratio $C(1 + |G|)$ also gives

$$\sup_{x\ge 0} Q_x(|W_1 - x| > t) \le Ce^{-ct}, \qquad t \ge 0.$$

A union bound and the entrance-time tail imply, for large fixed C,

$$Q_0\left(\nu_n > (\log n)^6 \text{ or } W_{\nu_n} > x_n + C\log n\right) = o(1).$$

For $x \in [x_n, x_n + C\log n]$,

$$\frac{x + b_n + B_h}{h(x)} = 1 + O((\log n)^{-1}), \qquad \frac{x + \sqrt{n} + B_h}{h(x)} n^{-D/2} = o(1) \quad (D > 4).$$

The Markov property at ν_n follows by conditioning on its integer values. For every bounded variable Ψ determined by the stopped path, $\Psi \mathbb{1}(\nu_n = k)$ is measurable with respect to $\sigma(W_0, \dots, W_k)$. Thus, for bounded measurable g and $r \in \mathbb{N}$, the ordinary Markov property gives

$$\begin{aligned} &\mathbb{E}_0^Q\left[\Psi \mathbb{1}(\nu_n = k) g(W_{k+1}, \dots, W_{k+r})\right] \\ &\quad = \mathbb{E}_0^Q\left[\Psi \mathbb{1}(\nu_n = k) \mathbb{E}_{W_k}^Q[g(W_1, \dots, W_r)]\right]. \end{aligned}$$

Summing over $k \ge 0$ then gives the conditional future law $Q_{W_{\nu_n}}$ at ν_n, which is finite almost surely by the preceding exit bound.

For ranks $j \ge (\log n)^{16}$, the entrance time, starting height, and path enlargement change the standardized heights by $o_{\mathbb{P}}(L_n^{-1/2})$. Indeed, covering $[0, n]$ by overlapping intervals of length 2δ and applying (SA-20) gives

$$\mathbb{P}\left(\sup_{\substack{0\le s,t\le n \\ |t-s|\le\delta}} |\mathcal{B}(t) - \mathcal{B}(s)| > x \right) \le C(1 + n/\delta) \exp\{-x^2/(C\delta)\}, \qquad 1 \le \delta \le n.$$

With $\delta = (\log n)^6$, the modulus is $O_{\mathbb{P}}((\log n)^{7/2})$. Couple Bessel processes by the norms of the same three-dimensional Brownian motion. Their starting-height error is at most $C(\log n)^2$. Thus, uniformly for $j \geq (\log n)^{16}$,

$$\frac{O_{\mathbb{P}}((\log n)^{7/2}) + O((\log n)^2)}{\sqrt{j}} = o_{\mathbb{P}}(L_n^{-1/2}), \qquad \log\left(1 + \frac{(\log n)^6}{j}\right) = o(1).$$

Integer sampling has the same bound. At heights $O(\sqrt{L_n})$, the squared-threshold error is $O(\sqrt{L_n})o_{\mathbb{P}}(L_n^{-1/2}) = o_{\mathbb{P}}(1)$. Hence (SA-43) applies separately to exceedance and no-exceedance events.

For three-dimensional Brownian motion $\mathbf{B}$, $e^{-s/2}\mathbf{B}(e^s)$ has three independent stationary O-U coordinates, each with covariance $e^{-|s-t|/2}$. Lemma SA-35, with $d = 3$, gives the maximum law from [3, Lemma 2.1, p. 673]. On a range of logarithmic length $\ell \log m + o(\log m)$, the CDF at squared threshold $v_m(u)$ therefore converges to

$$\exp\{-c_* \ell e^{-u}\}, \qquad c_* = \Gamma(3/2)^{-1}.$$

Squaring the lemma's threshold, more explicitly, gives

$$\left\{\frac{2L_m + \frac{3}{2}\log L_m - \log\Gamma(3/2) + z}{\sqrt{2L_m}}\right\}^2 = 2L_m + 3\log L_m - 2\log\Gamma(3/2) + 2z + o(1).$$

Fix $c_J, C_J > 0$. For any integer interval $J_m = [a_m, b_m] \cap \mathbb{N}$ satisfying $a_m \geq c_J(\log m)^{16}$, $b_m \leq m^{C_J}$, and $\log(b_m/a_m)/\log m \to \ell > 0$, apply (SA-43) with horizon $\lceil b_m \rceil$. The standardized-height error is $O_{\mathbb{P}}((\log m)^{7/2}/\sqrt{a_m}) = o_{\mathbb{P}}(L_m^{-1/2})$ uniformly on J_m. The comparison therefore gives

$$\limsup_{m\to\infty} Q_0\left(\max_{j\in J_m} \frac{W_j^2}{j} \leq v_m(u)\right) \leq \exp\{-c_*\ell e^{-u}\},$$
$$\limsup_{m\to\infty} Q_0\left(\max_{j\in J_m} \frac{W_j^2}{j} > v_m(u)\right) \leq 1 - \exp\{-c_*\ell e^{-u}\} \leq c_*\ell e^{-u}.$$

Using $u \pm \eta$ and then $\eta \downarrow 0$ absorbs the $o(1)$ threshold perturbations.

Apply (SA-37) once more. On each dyadic block $[b, 2b] \subset [1, m/2]$, (SA-38) gives

$$\mathbb{P}\left(\max_{b\leq j\leq 2b} E_m(j) > c\sqrt{bL_m}\right) \leq C_r L_m^{-r}.$$

Each of $[1, (\log m)^{16}]$ and $[m/(\log m)^4, m/2]$ uses $O(L_m)$ blocks, so its total contribution is $O(L_m^{1-r}) = o(1)$ for $r > 1$. On the retained ranks,

$$\frac{A_m(j)^2}{j} = 1 + O((\log m)^{-4}), \qquad v_m(u)(\log m)^{-4} = o(1).$$

The logarithmic lengths are, in turn,

$$\log\frac{m^a}{(\log m)^{16}} = a\log m + o(\log m) \qquad \text{and} \qquad \log\frac{m/(\log m)^4}{m^\beta} = (1-\beta)\log m + o(\log m).$$

This proves (SA-40) and (SA-41), also on the reversed half. On both whole halves the bound is $Ce^{-u} + o(1)$. For $A > \sqrt{2}$, $v_m(u) < A^2 L_m$ eventually, so

$$\limsup_{m\to\infty} \mathbb{P}\left(\max_{1\leq j<m} \frac{E_m(j)^2}{A_m(j)^2} > A^2 L_m\right) \leq Ce^{-u} \to 0 \quad (u \to \infty).$$

For (SA-42), apply the no-exceedance bound through $H/(\log H)^4$, using

$$\log\frac{H/(\log H)^4}{m^\vartheta} = (\gamma - \vartheta)\log m + o(\log m) \qquad \text{and} \qquad \log\frac{H/(\log H)^4}{(\log H)^{16}} = \gamma\log m + o(\log m).$$

□

Localization of Gaussian near-maximizers. We now bound the probability that an excursion stays below a winning band, as a step toward localizing nearly maximizing cuts. Put

$$D_m(z) = \mathbb{P}\big(E_m(j) \le z A_m(j) \text{ for all } 1 \le j < m \big).$$

There are absolute $c, C, z_0 > 0$ such that

$$D_m(z) \ge c \exp\{-C(1+\log m) z^3 e^{-z^2/2}\}, \qquad m \ge 2, \quad z \ge z_0. \tag{SA-44}$$

To prove this bound, let D be the discrete excursion law and C the integer skeleton of a continuous Brownian excursion of duration m. Using the killed Brownian kernel $k_t(x,y) = \phi_t(y-x) - \phi_t(y+x)$ and taking the two endpoints to zero gives the exact density ratio

$$\frac{d\mathsf{C}}{d\mathsf{D}}(x_1, \dots, x_{m-1}) = 2 x_1 x_{m-1} \prod_{j=1}^{m-2} (1 - e^{-2 x_j x_{j+1}}).$$

Indeed, (SA-36) and

$$\frac{k_1(\epsilon, x)}{\epsilon} \to 2x\phi_1(x), \qquad \frac{k_m(\epsilon, \eta)}{\epsilon \eta} \to \frac{2\phi_m(0)}{m}$$

give the two end factors. If $x_1, x_{m-1} \in [1,2]$, then $d\mathsf{C}/d\mathsf{D} \le 2 x_1 x_{m-1} \le 8$.

The continuous excursion, in turn, is a three-dimensional Bessel bridge from zero to zero and has the law of the Euclidean norm of three independent Brownian bridges [21, Section 7.1, Eq. (92)]. For one such bridge B, put $s(t) = \log\{t/(m-t)\}$. For $0 < t \le u < m$,

$$\mathbb{E}\left[\frac{B(t)}{A_m(t)} \frac{B(u)}{A_m(u)} \right] = \frac{t - tu/m}{A_m(t) A_m(u)} = \sqrt{\frac{t(m-u)}{u(m-t)}} = e^{-\{s(u) - s(t)\}/2}.$$

Thus dividing each coordinate by $A_m(t)$ and using the time change $s = s(t)$ gives a stationary standard three-dimensional O-U process. The times $1 \le t \le m-1$ therefore correspond to an interval of length $T = 2\log(m-1)$. For $m \ge 4$, let $\mathcal{A} = \{\mathbf{x} \in \mathbb{R}^3 : 1.2 \le \|\mathbf{x}\| \le 1.8\}$. O-U endpoints in $\mathcal{A}$ give excursion heights at times $1, m-1$ in $[1,2]$. If $P_T^{(z)}$ is the O-U semigroup killed on leaving the ball of radius z, and γ_3 is standard three-dimensional Gaussian measure, it follows that

$$D_m(z) \ge \tfrac{1}{8} \langle \mathbb{1}(\cdot \in \mathcal{A}), P_T^{(z)} \mathbb{1}(\cdot \in \mathcal{A}) \rangle_{L^2(\gamma_3)}.$$

The Dirichlet form is $\frac{1}{2}\int \|\nabla f\|^2 \, d\gamma_3$. For $\psi_z(\mathbf{x}) = \min\{1, z(z - \|\mathbf{x}\|)_+\}$, the Rayleigh quotient [5, Section 2.1, Eq. (2.9)] gives

$$\lambda_z \le \frac{\frac{1}{2}\int \|\nabla \psi_z\|^2 \, d\gamma_3}{\int \psi_z^2 \, d\gamma_3} \le C z^2 \gamma_3\big(\{\mathbf{x} \in \mathbb{R}^3 : z - 1/z < \|\mathbf{x}\| < z\}\big) \le C z^3 e^{-z^2/2}.$$

Here $\int \psi_z^2 \, d\gamma_3$ is bounded below for large z. The first eigenfunction can be chosen nonnegative [5, Lemma 2.1]. Denote it by f_z, normalize it in $L^2(\gamma_3)$, and extend it by zero. The Gaussian Poincaré inequality gives, in turn,

$$1 - \left(\int f_z \, d\gamma_3 \right)^2 \le 2\lambda_z.$$

Indeed, the unrestricted O-U semigroup satisfies

$$P_t f(\mathbf{x}) = \int f\big(e^{-t/2}\mathbf{x} + \sqrt{1 - e^{-t}}\,\mathbf{y}\big) \, d\gamma_3(\mathbf{y}), \qquad \nabla P_t f = e^{-t/2} P_t \nabla f, \qquad -\frac{d}{dt}\int (P_t f)^2 \, d\gamma_3 = \int \|\nabla P_t f\|^2 \, d\gamma_3.$$

These follow from the Gaussian transition law, differentiation, and Gaussian integration by parts. Integrating and applying Jensen then gives

$$\mathbb{V}_{\gamma_3}[f] = \int_0^\infty \int \|\nabla P_t f\|^2 \, d\gamma_3 \, dt \le \int_0^\infty e^{-t} \, dt \int \|\nabla f\|^2 \, d\gamma_3.$$

The inequality extends to $H^1(\gamma_3)$ by smooth approximation, so it applies to the zero extension of f_z. Since $f_z \ge 0$ and $\|f_z\|_{L^2(\gamma_3)} = 1$,

$$\|f_z - 1\|_{L^2(\gamma_3)}^2 = 2\left(1 - \int f_z \, d\gamma_3\right) \le 2\{1 - \sqrt{1 - 2\lambda_z}\} \to 0,$$

$$\int_{\mathcal{A}} f_z \, d\gamma_3 \ge \gamma_3(\mathcal{A}) - \sqrt{\gamma_3(\mathcal{A})}\, \|f_z - 1\|_{L^2(\gamma_3)} \ge c > 0$$

for all $z \geq z_0$, increasing z_0 if necessary. The Dirichlet operator $-\frac{1}{2}(\Delta - \mathbf{x}\cdot\nabla)$ has a discrete self-adjoint spectrum [5, Section 2.1], with eigenvalues one half those in that reference. Its orthonormal eigenfunctions satisfy

$$-\tfrac{1}{2}(\Delta - \mathbf{x}\cdot\nabla) f_{z,j} = \lambda_{z,j} f_{z,j}, \qquad f_{z,j}|_{\partial B(0,z)} = 0, \qquad (f_{z,1}, \lambda_{z,1}) = (f_z, \lambda_z).$$

The boundary regularity argument in Lemma SA-41 permits stopped Itô [21, Section 5.3, p. 26]. The zero boundary values give $P_T^{(z)} f_{z,j} = e^{-\lambda_{z,j} T} f_{z,j}$, so

$$\langle \mathbb{1}(\cdot \in \mathcal{A}), P_T^{(z)} \mathbb{1}(\cdot \in \mathcal{A})\rangle = \sum_{j\geq 1} e^{-\lambda_{z,j} T} |\langle \mathbb{1}(\cdot \in \mathcal{A}), f_{z,j}\rangle|^2 \geq c e^{-\lambda_z T}.$$

This proves (SA-44). Decreasing c also covers $m = 2, 3$.

Lemma SA-46 (Global Gaussian near-maximizer bound)**.** *Let X be a centered Gaussian corridor bridge of length m under a scalar history as in Lemma SA-43. Reveal its winning rank k and score $z \geq z_0$, and put $q = \min\{k, m-k\}$. For $E \geq 0$ and an integer $r \geq 1$, the conditional probability that some internal rank j satisfies*

$$\min\{j, m-j, |j-k|\} > r, \qquad zA_m(j) - |X(j)| \leq 2E\left(1 + \frac{A_m(j)}{A_m(k)}\right)$$

is at most

$$C\exp\{C(1+\log m) z^3 e^{-z^2/2}\}\left\{\frac{1+E}{\sqrt{r}} + \frac{E(1+\log m)}{\sqrt{q}}\right\}.$$

The constants do not depend on the feasible history or on k, z, E, r. The same bound holds after conditioning on the sign of the winning height, and the right side is allowed to exceed one.

Proof. Condition also on $\varsigma = \operatorname{sgn} X(k)$, and reflect both the path and its inherited walls when $\varsigma = -1$. The winning height is now positive. Put $b(j) = zA_m(j)$. After revealing this winner the effective walls on either child are $\ell_* = \max\{\ell, -b\}$ and $u_* = \min\{u, b\}$, so $\ell_* \leq 0 \leq u_*$. The gaps $\Delta_+ = b - X$ and $\Delta_- = b + X$ have nonnegative endpoints. Write their lower and upper walls, respectively, as

$$\ell_+ = b - u_* \geq 0, \quad u_+ = b - \ell_* \geq b, \qquad \ell_- = b + \ell_* \geq 0, \quad u_- = b + u_* \geq b.$$

On a child with endpoints j_0, j_1, their densities, up to endpoint-dependent constants, are

$$\exp\left\{-\frac{1}{2}\sum_{i=j_0+1}^{j_1}(g_i - g_{i-1})^2 - \sum_{i=j_0+1}^{j_1-1}(b(i+1) - 2b(i) + b(i-1))g_i\right\}\prod_{i=j_0+1}^{j_1-1}\mathbb{1}(\ell_\pm(i) \leq g_i \leq u_\pm(i)).$$

Expanding the squared Gaussian increments and summing the cross term by parts gives this density. Since b is concave, all linear fields are nonnegative. By Lemma SA-44, each gap is therefore bounded below in stochastic order, separately, by the zero-endpoint excursion on that child conditioned to stay below the restriction of b.

If the child's length is s, this restricted parent band dominates zA_s. On the left child this follows from $j(1-j/m) \geq j(1-j/k)$. On the right child, with local rank i and $s = m - k$, the exact difference of squared bands divided by z^2 is

$$(k+i)(1-(k+i)/m) - i(1-i/s) = \frac{k(s-i)^2}{ms} \geq 0.$$

Thus for any decreasing gap event $\mathcal{E}$,

$$\mathbb{P}\,(\,\Delta_\pm \in \mathcal{E} \,|\, \mathcal{H},\, k,\, z,\, \varsigma\,) \leq \frac{\mathbb{P}(E_s \in \mathcal{E})}{D_s(z)}. \tag{SA-45}$$

The bound, in particular, does not divide by the probability of the accumulated ancestral corridor.

Reverse the interval so that $k = q \leq m/2$. The band ratios satisfy

$$\frac{A_m(j)}{A_m(k)} \leq \begin{cases} \sqrt{2}, & j < k, \\ \sqrt{2(1+d/q)}, & d = j-k \leq (m-k)/2, \\ \sqrt{2d/q}, & d = m-j \leq (m-k)/2. \end{cases}$$

On each half-child, the event therefore requires a gap below $CE(1+\sqrt{d/q})$ at a distance $d > r$. For a dyadic block $[t, 2t]$, (SA-45) and (SA-39) bound its probability by

$$D_s(z)^{-1}\frac{C\{1 + E(1+\sqrt{2t/q})\}}{\sqrt{t}}.$$

Summing over $t = 2^j r \le m$ then gives

$$\sum_{\substack{j\ge 0\\ 2^j r\le m}} \left\{ \frac{1+E}{\sqrt{2^j r}} + \frac{E}{\sqrt{q}} \right\} \le C \left\{ \frac{1+E}{\sqrt{r}} + \frac{E(1+\log m)}{\sqrt{q}} \right\}.$$

Sum over the two gaps and four half-child ranges, and use (SA-44) for $2 \le s \le m$. A child of length one has no internal rank. This proves the bound for each ς and hence after averaging over its conditional law. No symmetry of the inherited corridor is used. □

For fixed y and all sufficiently large m, uniformly over $2 \le m' \le m$, equation (SA-44) gives

$$\begin{aligned} z^2 \ge 2L_m + \log L_m - y \quad &\Longrightarrow \quad D_{m'}(z)^{-1} \le C(\log m)^C, \\ z^2 \ge v_m(-y) \quad &\Longrightarrow \quad D_{m'}(z)^{-1} \le C. \end{aligned}$$

For any partial-sum perturbation bounded by E, write $\widehat{X}$ for its centered bridge. Then $\|\widehat{X} - X\|_\infty \le 2E$. If j maximizes its standardized score and k maximizes the Gaussian score,

$$\frac{|X(j)| + 2E}{A_m(j)} \ge \frac{|\widehat{X}(j)|}{A_m(j)} \ge \frac{|\widehat{X}(k)|}{A_m(k)} \ge z - \frac{2E}{A_m(k)}.$$

Hence, simultaneously for every such perturbation,

$$zA_m(j) - |X(j)| \le 2E \left(1 + \frac{A_m(j)}{A_m(k)} \right).$$

No independence is required.

Transfer from Gaussian bridges to the construction tree.

Lemma SA-47 (Small node at log–log depth)**.** *There are constants $B, C_0, q_{\mathrm{node}} > 0$ such that, for all sufficiently large N, the maximal scalar construction tree grown without a depth cap has, with probability at least q_{node}, a node containing at most $2(\log N)^B$ observations at depth at most $\lceil C_0 \log\log N \rceil$. The event and the selected node can be chosen using only $(Z_i)_{i=1}^N$ and auxiliary randomness independent of the numerical covariate order statistics.*

Proof. By the Komlós–Major–Tusnády approximation, as stated in [2, Theorem 2.6.2, p. 107] and [23, Lemma 4.2], on an extension of the probability space there are independent standard normal variables $G_1, \dots, G_N$ such that, with $T_j = \sum_{i=1}^j G_i$ and $T_0 = 0$,

$$\mathcal{K}_N = \left\{ \max_{0\le j\le N} |S_j - T_j| \le \Delta_N \right\}, \qquad \Delta_N = C_{\mathrm{KMT}} \log N, \qquad \mathbb{P}(\mathcal{K}_N^c) = o(1). \tag{SA-46}$$

The exponential moment assumption gives

$$\lim_{\lambda \downarrow 0} \lambda \mathbb{E}\left[|Z_i|^3 e^{\lambda |Z_i|} \right] = 0.$$

Choose $\lambda > 0$ with $\lambda \mathbb{E}[|Z_i|^3 e^{\lambda|Z_i|}] < \mathbb{E}[Z_i^2] = 1$. Sakhanenko's exponential inequality in [23, Lemma 4.2] and Markov's inequality then give, for a constant $c_{\mathrm{KMT}} > 0$,

$$\mathbb{P}(\mathcal{K}_N^c) \le (1 + \lambda\sqrt{N}) N^{-c_{\mathrm{KMT}} \lambda C_{\mathrm{KMT}}} = o(1)$$

when the fixed constant C_{KMT} satisfies $c_{\mathrm{KMT}} \lambda C_{\mathrm{KMT}} > 1$. No density assumption is needed. Realize the regular conditional law of $(G_i)_{i=1}^N$ given $(Z_i)_{i=1}^N$ [18, Theorem 4.1.18] by successive conditional quantiles [18, proof of Theorem 1.2.2], using auxiliary uniforms independent of the construction covariates and the estimation fold. Writing $U_{(1)}, \dots, U_{(N)}$ for the transformed construction order statistics, this gives

$$((Z_i, G_i))_{i=1}^N \perp\!\!\!\perp (U_{(1)}, \dots, U_{(N)}).$$

The G_i are mutually independent, but may depend on (Z_i). Gaussian histories reveal only Gaussian quantities. Coupling success $\mathcal{K}_N$ is instead imposed only on the final event.

We retain the notation A_m, L_m, and v_m from the Gaussian lemmas. In intermediate inequalities, C and c denote positive constants that may change from line to line but do not depend on N, the revealed history, or the interval's position.

Gaussian intervals and their revealed boundary values. Represent an interval of observations by $I = (s, s+m]$, meaning the ranks $s+1, \dots, s+m$. Its mean and centered Gaussian bridge are

$$\mu_I = \frac{T_{s+m} - T_s}{m}, \qquad X_I(t) = T_{s+t} - T_s - \mu_I t, \quad t \in \{0, \dots, m\}.$$

The Gaussian history determines the endpoint sums of I and the increments in the adjacent intervals $(s-\ell_{\mathrm{L}}, s]$ and $(s+m, s+m+\ell_{\mathrm{R}}]$, where $0 \le \ell_{\mathrm{L}}, \ell_{\mathrm{R}} \le R$ and R is the interval's recorded integer radius. All endpoints are integers, as are the offsets i, i' below, and all intervals lie in $(0, N]$. The adjacent increments have already been revealed, belong to no unexplored Gaussian interval, and satisfy

$$\max_{0 \le i \le \ell_{\mathrm{L}}} |T_s - T_{s-i} - \mu_I i| \le P\sqrt{R}, \qquad \text{and} \qquad \max_{0 \le i' \le \ell_{\mathrm{R}}} |T_{s+m+i'} - T_{s+m} - \mu_I i'| \le P\sqrt{R}. \tag{SA-47}$$

The fixed constant P is chosen below. An actual construction node matched to this Gaussian interval is allowed to be any

$$J = (s - i, s + m + i'], \qquad 0 \le i \le \ell_{\mathrm{L}}, \quad 0 \le i' \le \ell_{\mathrm{R}}. \tag{SA-48}$$

The comparison with the actual tree will hold simultaneously for every such J and every sequence satisfying (SA-46).

Conditional on the Gaussian history, the bridge X_I is restricted between a convex lower wall and a concave upper wall, denoted by $\mathcal{W}_I^-$ and $\mathcal{W}_I^+$. The bridge vanishes at both endpoints, and $\mathcal{W}_I^-(t) \le 0 \le \mathcal{W}_I^+(t)$. The intervals to which a complete two-split step is applied satisfy, after reversing the order and changing sign if necessary,

$$\begin{aligned} &\mathcal{W}_I^-(t) \le -w(1 - t/m), \qquad 0 \le \mathcal{W}_I^+(t) \le \delta + Dt, \qquad 0 \le t \le m, \\ &w \ge c_{\mathrm{w}}\sqrt{m}, \qquad 0 \le D \le C_{\mathrm{D}}\sqrt{L_m/m}, \qquad 0 \le \delta \le P\sqrt{R}. \end{aligned} \tag{SA-49}$$

Here $c_{\mathrm{w}}, C_{\mathrm{D}} > 0$ are fixed, and $A > \sqrt{2}$ is a fixed score cap. Reversing the order and changing the sign serve only to express the wall bounds and do not alter the rank interval or the revealed increments. The intermediate larger interval in a two-split step, by contrast, satisfies the weaker bounds

$$w \ge c_{\mathrm{w}} m^{a/2}, \qquad D \le C_{\mathrm{D}}\sqrt{L_m}\, m^{-a/2}, \tag{SA-50}$$

for a fixed $a > 0$. The other parts of (SA-47)–(SA-49) are unchanged.

Splitting from a wall with a small positive intercept. Lemmas SA-44 and SA-45 together give the following bound. Suppose a centered corridor bridge of length m satisfies

$$\mathcal{W}^-(t) \le -w(1 - t/m), \qquad 0 \le \mathcal{W}^+(t) \le \delta + Dt, \qquad w \ge cm^{\kappa}, \quad 0 \le D \le C\sqrt{L_m}\, m^{-\kappa},$$

where $0 < \kappa \le 1/2$ is fixed. Denote the conditioning Gaussian sigma-field by $\mathcal{H}$. Put $H = \lfloor m^{3\kappa/2} \rfloor$ and $J_m = \lceil m^{\vartheta} \rceil$, where $0 < \vartheta < 3\kappa/2$. Fix a deterministic sequence $\omega_m \downarrow 0$. For histories satisfying

$$\delta\sqrt{L_m}/\sqrt{J_m} \le \omega_m, \tag{SA-51}$$

uniformly over this class,

$$\mathbb{P}\left(\max_{J_m \le j \le H/(\log H)^4} \frac{|X(j)|^2}{A_m(j)^2} \le v_m(u) \,\middle|\, \mathcal{H} \right) \le C\exp\{-c(3\kappa/2 - \vartheta)e^{-u}\} + o(1). \tag{SA-52}$$

We first control the change caused by removing the remote wall. The Gaussian prefix bound gives $\mathbb{P}(X(H) < -w/4 \mid \mathcal{H}) \le CH/w^2 = o(1)$. Condition further on

$$\mathcal{H} \vee \sigma\{X(j) : H \le j \le m\}, \qquad X(H) \ge -w/4.$$

Let ν be the conditional prefix law with its lower wall removed. Decrease its upper wall to zero and its last endpoint to $-w/4$, and denote the resulting smaller bridge by Y. Its chord is then $t \mapsto -wt/(4H)$. The signed excursion comparison and the prefix bound on the two halves give

$$\mathbb{P}\left(\min_{0 \le j \le H} Y(j) < -w/2 \right) \le \mathbb{P}\left(\max_{0 \le j \le H} E_H(j) > w/4 \right) \le \frac{CH}{w^2}.$$

Since $H = o(m)$, the original lower wall is at most $-w/2$ throughout $[0, H]$. Put $\mathcal{S} = \{x(j) \ge \mathcal{W}^-(j),\ 1 \le j < H\}$. Stochastic ordering gives $\nu(\mathcal{S}^c) \le CH/w^2$, and the conditional prefix law with both walls is $\nu(\cdot \mid \mathcal{S})$. Therefore

$$\sup_{\mathcal{E} \text{ measurable}} \left| \nu(\mathcal{E} \mid \mathcal{S}) - \nu(\mathcal{E}) \right| = \nu(\mathcal{S}^c) \le \frac{CH}{w^2} = o(1).$$

The bound is uniform over the revealed suffixes under consideration.

With the lower wall removed, increase the upper wall to $\delta + Dt$ and the last endpoint to $\delta + DH$. The gap between this affine wall and the resulting larger bridge is a nonnegative Gaussian bridge from δ to zero. Increasing the initial endpoint from zero to δ shows that this gap stochastically dominates the excursion E_H of length H. Apart from the preceding total variation error, the original bridge is therefore stochastically bounded above by

$$\delta + Dj - E_H(j).$$

Put $\mathcal{J}_m = \{J_m, \dots, \lfloor H/(\log H)^4 \rfloor\}$. The two-sided no-exceedance event is contained in the increasing event $\{X(j) \geq -\sqrt{v_m(u)}A_m(j) \text{ for all } j \in \mathcal{J}_m\}$. Apply the preceding stochastic comparison to this latter event and average over the revealed suffix. Including the probability of $X(H) < -w/4$ gives

$$\mathbb{P}\left(|X(j)| \leq \sqrt{v_m(u)}A_m(j) \text{ for all } j \in \mathcal{J}_m \,\middle|\, \mathcal{H} \right) \leq \mathbb{P}\left(E_H(j) \leq \delta + Dj + \sqrt{v_m(u)}A_m(j) \text{ for all } j \in \mathcal{J}_m \right) + \frac{CH}{w^2}.$$

On $\mathcal{J}_m$, (SA-51) gives

$$\sup_{j \in \mathcal{J}_m} \frac{\delta + Dj}{\sqrt{j}} \leq \frac{\delta}{\sqrt{J_m}} + D\sqrt{H} \leq \frac{\omega_m}{\sqrt{L_m}} + C\sqrt{L_m}\, m^{-\kappa/4} = o(L_m^{-1/2}).$$

Since $A_m(j) \leq \sqrt{j}$, the squared threshold changes by $o(1)$. Lemma SA-45 therefore proves (SA-52). Its upper-exceedance and score-cap bounds require only a centered convex–concave corridor containing zero.

Comparison with every enlarged actual node. Fix $I = (s, s+m]$ and let k be its Gaussian maximizing rank. Write $z = |X_I(k)|/A_m(k)$ and $q = \min\{k, m-k\}$. For an enlarged node J in (SA-48), let $m_J = m + i + i'$ and define

$$B_J(t) = \sqrt{\frac{(t+i)(m+i'-t)}{m_J}}.$$

The identities

$$\frac{1}{B_J(t)^2} = \frac{1}{t+i} + \frac{1}{m+i'-t} \leq \frac{1}{t} + \frac{1}{m-t} = \frac{1}{A_m(t)^2} \qquad \text{and} \qquad \frac{B_J(k)}{A_m(k)} \leq 1 + \frac{R}{q}$$

hold because

$$\frac{B_J(k)^2}{A_m(k)^2} = \frac{m}{m_J}\left(1 + \frac{i}{k}\right)\left(1 + \frac{i'}{m-k}\right) \leq \left(1 + \frac{R}{q}\right)^2.$$

Write $\Pi_{\mathrm{L}}(i) = T_s - T_{s-i} - \mu_I i$ and $\Pi_{\mathrm{R}}(i') = T_{s+m+i'} - T_{s+m} - \mu_I i'$. Centering the Gaussian partial sum over J instead of I changes its value at rank t of I by

$$\Pi_{\mathrm{L}}(i) - \frac{t+i}{m_J}(\Pi_{\mathrm{L}}(i) + \Pi_{\mathrm{R}}(i')).$$

Its absolute value is at most $|\Pi_{\mathrm{L}}(i)| + |\Pi_{\mathrm{R}}(i')| \leq 2P\sqrt{R}$. Replacing T by S, in turn, changes an interval-centered partial sum by at most $4\Delta_N$ on $\mathcal{K}_N$. Thus the total numerator error is at most

$$\Delta_* = 2P\sqrt{R} + 4\Delta_N.$$

The actual score at the valid candidate k is therefore at least

$$\frac{zA_m(k) - \Delta_*}{B_J(k)} \geq z - zR/q - \Delta_*/A_m(k).$$

In the parameter range chosen below this lower bound is positive. Consequently every actual maximizing cut lying at an internal rank j of I satisfies

$$zA_m(j) - |X_I(j)| \leq \Delta_* + \left\{ \frac{\Delta_*}{A_m(k)} + \frac{zR}{q} \right\} A_m(j). \qquad \text{[SA-53]}$$

Indeed, maximality at j and $B_J(j) \geq A_m(j)$ give

$$\frac{|X_I(j)| + \Delta_*}{A_m(j)} \geq \frac{|X_I(j)| + \Delta_*}{B_J(j)} \geq z - \frac{zR}{q} - \frac{\Delta_*}{A_m(k)}.$$

Since $A_m(k) \leq \sqrt{q}$, the right-hand side of (SA-53) is bounded by $\Delta^\dagger\{1 + A_m(j)/A_m(k)\}$, where

$$\Delta^\dagger = \Delta_* + zR/\sqrt{q}. \qquad \text{[SA-54]}$$

Conditional on the Gaussian history, the winning rank, and its signed height, apply Lemma SA-46 with $2E = \Delta^\dagger$ to this Gaussian near-band event. Except on a Gaussian event with conditional probability at most

$$C \exp\{C(1+\log m) z^3 e^{-z^2/2}\} \left\{ \frac{1+\Delta^\dagger}{\sqrt{r}} + \frac{\Delta^\dagger (1+\log m)}{\sqrt{q}} \right\}, \tag{SA-55}$$

every actual maximizing cut either creates a child with at most $R + r$ observations or lies within r ranks of k. Cuts outside I and at or near either endpoint of I are included. The event holds simultaneously for all allowed padding lengths and all partial-sum perturbations in (SA-46), regardless of any dependence between the coupling error and the Gaussian bridge. It also covers all maximizing actual ranks when there are ties.

For the nonroot steps we take

$$r = FR/2, \qquad R' = FR,$$

where $F \geq 4$ is a fixed even integer. Recorded radii are integers. The cutoff choices below ensure, uniformly,

$$\Delta_N \leq \sqrt{R}, \qquad r = o(q), \qquad z\sqrt{R/q} = o(1), \qquad \Delta^\dagger \leq C(P+1)\sqrt{R}.$$

For fixed u and $z^2 \geq v_m(u)$, $(1+\log m) z^3 e^{-z^2/2} = O(e^{-u})$, since $z^3 e^{-z^2/2}$ is eventually decreasing. Thus (SA-55) is at most

$$C_u (P+1) \left\{ F^{-1/2} + \sqrt{R/q}(1+\log m) \right\}.$$

Choose F large. The second term, meanwhile, is uniformly $o(1)$ above the cutoff.

Trimming and the next conditional Gaussian state. Reveal the Gaussian increments in the middle interval $(s+k-r, s+k+r]$, and retain the two intervals

$$(s, s+k-r], \qquad (s+k+r, s+m].$$

On localization near k, each is contained in the corresponding actual child. The enlarged endpoints extend it by at most $2r = R'$ at its new edge and at most R at its old edge.

Conditional on the Gaussian winner and its signed height, the two untrimmed child bridges are independent corridor bridges. On each child, require the absolute centered partial sums, read from the cutting endpoint, to be at most $H_0\sqrt{r}$ through the first r ranks. The prefix inequality then bounds the conditional probability of failure by C/H_0^2 when $r \leq q/2$. These inequalities involve increments revealed in the middle interval, so they are measurable from the subsequent Gaussian history.

To check the walls and the complete padding bound, change sign so that $X_I(k) = b(k)$, where $b(t) = zA_m(t)$. For the left trimmed interval, put $t = k - r$ and write

$$X_I(t) = b(t) - \delta', \qquad 0 \leq \delta' \leq H_0\sqrt{r} + Czr/\sqrt{q}. \tag{SA-56}$$

The winning band and the revealed prefix bound give, for $r \leq q/2$,

$$0 \leq \delta' = b(t) - \frac{t}{k} b(k) - \left\{ X_I(t) - \frac{t}{k} X_I(k) \right\} \leq \frac{Czr}{\sqrt{q}} + H_0\sqrt{r},$$

$$\frac{zr}{\sqrt{q}} = o(\sqrt{r}), \qquad \delta' \leq P\sqrt{R'}, \qquad X_I(t) \geq cz\sqrt{q}.$$

The right interval follows by reversal. Here P is sufficiently large and fixed, and the last inequalities hold for large N.

Orient the new bridge away from its new cutting endpoint. Every old lower wall, as well as the negative winning band, lies below the old parent chord. Subtracting the new chord consequently gives the lower wedge $-X_I(t)(1-j/t)$. The tangent to its positive winning wall has intercept δ' and slope at most

$$-b'(t) + \frac{X_I(t)}{t} = \frac{z\sqrt{m}}{2\sqrt{t(m-t)}} - \frac{\delta'}{t} \leq Cz/\sqrt{q}. \tag{SA-57}$$

The lower wall, moreover, remains convex, the upper wall remains concave, and the new endpoint chord lies between them. The smaller trimmed interval has length asymptotic to q and hence satisfies (SA-49) with constants independent of the number of preceding steps. For the larger trimmed interval, $w \geq c\sqrt{q}$ and $D \leq Cz/\sqrt{q}$, which gives (SA-50) when $q \geq m^a$.

For an original child $\nu \in \{\mathtt{L}, \mathtt{R}\}$ of size n_ν, write Y_ν for its centered bridge read away from the cut and μ'_ν for its trimmed mean. Since the removed sum is $r\mu_\nu + Y_\nu(r)$,

$$\mu'_\nu - \mu_\nu = -\frac{Y_\nu(r)}{n_\nu - r}, \qquad |\mu'_\nu - \mu_\nu| \leq \frac{H_0\sqrt{r}}{n_\nu - r}.$$

The newly revealed pad has length $2r$, not merely r. The slopes of the two original child chords differ in absolute value by

$$\frac{z}{\sqrt{q(1-q/m)}} = O(z/\sqrt{q}).$$

Define the two orientations of the centered pad sums by

$$\widetilde{\Pi}_{\mathtt{L}}(i) = \sum_{j=1}^{i} (G_{s+k-r+j} - \mu'_{\mathtt{L}}), \qquad \widetilde{\Pi}_{\mathtt{R}}(i) = \sum_{j=1}^{i} (G_{s+k+r-j+1} - \mu'_{\mathtt{R}}).$$

Expressing each within-child segment as a difference of two revealed centered prefixes yields

$$\begin{aligned} \max_{\nu\in\{\mathtt{L},\mathtt{R}\}} \max_{0\le i\le 2r} |\widetilde{\Pi}_\nu(i)| &\le 4H_0\sqrt{r} + 2r|\mu_{\mathtt{L}} - \mu_{\mathtt{R}}| + 2r \max_{\nu\in\{\mathtt{L},\mathtt{R}\}} |\mu'_\nu - \mu_\nu| \\ &\le 4H_0\sqrt{r} + Czr/\sqrt{q} + CH_0 r^{3/2}/q \le P\sqrt{R'}. \end{aligned} \tag{SA-58}$$

Choosing, for example, $P \ge 8H_0$ makes (SA-58) hold for large N independently of F, since $\sqrt{R'} = \sqrt{2r}$.

For an inherited outside pad, recentering gives the bound

$$P\sqrt{R} + \frac{CRz}{\sqrt{q}} + \frac{CH_0 R\sqrt{r}}{q} = P\sqrt{R} + o(\sqrt{R}) \le P\sqrt{R'}.$$

Together with (SA-58), this proves (SA-47) on both entire pads. Given the cut, signed height, and middle-pad increments, the shared padding is fixed. Lemma SA-43 therefore gives independent remaining interiors.

A two-split step. Fix $\varrho \in (0, 1/4)$. We now choose the constants so that a complete step starting from (SA-49) has conditional failure probability at most ϱ, uniformly in the history. Let $\gamma = 3/4$, and let $\eta > 0$ be small enough compared with ϱ. First choose $\lambda_1 = e^{-u_1}$ so large that

$$C\exp\{-c(\gamma/2)\lambda_1\} < \eta.$$

Choose $0 < a < \gamma/2$ so small that $Ca\lambda_1 < \eta$. Next choose $\lambda_2 = e^{-u_2}$ so large that

$$C\exp\{-c(\gamma a/2)\lambda_2\} < \eta,$$

and then choose $a_2 > 0$ so small that

$$2a_2 < \min\{a, \gamma a/2\}, \qquad C(2a_2)\lambda_2 < \eta.$$

Finally choose $\beta < 1$, with $\beta > \max\{a, 1/2\}$, sufficiently close to one that $C(1-\beta)\max\{\lambda_1, \lambda_2\} < \eta$. The constants C, c in these choices include the fixed multiplicities from the two signs and two endpoints in the excursion estimates.

For the first cut, take $\kappa = 1/2$ and $\vartheta = a$ in (SA-52). Together with the two-endpoint, central, and score-cap bounds in Lemma SA-45, transferred by Lemma SA-44, this gives

$$\mathbb{P}\left(\{m^a \le q \le m^\beta,\ v_m(u_1) < z^2 \le A^2 L_m\}^c \,\middle|\, \mathcal{H}\right) \le Ce^{-c(\gamma-a)\lambda_1} + Ca\lambda_1 + C(1-\beta)\lambda_1 + o(1) \le C\eta + o(1).$$

Trim and reveal, leaving the smaller interval unexplored. The larger trimmed interval then has $m'/m \to 1$ and satisfies (SA-50). Conditional on these revelations, apply the same calculation with

$$\kappa = a/2, \qquad \vartheta = 2a_2, \qquad \gamma a - 2a_2 \ge \gamma a/2.$$

Its smaller count and score therefore satisfy

$$(m')^{2a_2} \le q \le (m')^\beta, \qquad v_{m'}(u_2) < z^2 \le A^2 L_{m'},$$

apart from conditional probability $Ce^{-c(\gamma a - 2a_2)\lambda_2} + C(2a_2)\lambda_2 + C(1-\beta)\lambda_2 + o(1) \le C\eta + o(1)$.

Choose H_0 to make the total probability of the two padding failures at most η, and take $P \ge 8H_0$. Then choose the even integer F so that the first terms of the two bounds (SA-55) sum to at most η. The second call, in turn, uses radius FR. Recording both output radii as F^2R only weakens the padding bounds already obtained and does not require revealing any further increments.

These estimates hold uniformly provided

$$RL_m/m^{a_2} \to 0 \tag{SA-59}$$

and the second terms of (SA-55) tend to zero. Condition (SA-59) ensures $r = o(m^{a_2})$. The deterministic cutoff bounds below give the sequence ω_m required in (SA-51) for both retained ranges. After trimming, the two smaller output sizes satisfy

$$m^a - O(R) \ge m^{a_2}, \qquad (m')^{2a_2} - O(FR) \ge m^{a_2}, \qquad \max\{m_{\mathrm{out},1}, m_{\mathrm{out},2}\} \le m^\beta.$$

Both obey (SA-47)–(SA-49). The second middle pad lies inside the larger interval and misses the waiting smaller interval. Their remaining interiors are disjoint and conditionally independent.

The two cuts and the middle-pad increments determine whether the scalar size, height, and padding requirements hold. They, however, need not determine the localization events, which we do not add to the history. For η sufficiently small, the conditional probability of failure of either the scalar requirements or localization is at most ϱ, uniformly above the cutoff chosen below. If both localizations hold, each output interval matches an actual descendant at most two splits below the incoming actual node, unless an earlier endpoint split has already produced an actual child with at most CF^2R observations. Such a child already satisfies the required size bound. When the intervals are matched, the Gaussian and actual smaller sides agree, since the smaller count is at most $m^\beta = o(m)$ and the rank uncertainty is $o(q)$.

The conditional failure recursion. Index the reference intervals by finite binary words v, with successors $v0$ and $v1$, and let $\mathcal{H}_v$ be the Gaussian sigma-field before the two-split step at v. At the first reference interval this history consists of the root total sum, maximizing rank, signed height, and revealed middle increments. At later intervals it contains only the revelations along the ancestral reference steps. Both children receive the history $\mathcal{H}_v^+$ defined below, and later revelations inside one child are not added to the other child's history. Let m_v and R_v denote the interval's length and recorded radius. A history is admissible if its root revelations satisfy $\mathcal{A}_{\mathrm{rt}} \cap \mathcal{P}_{\mathrm{rt}}$ and every ancestral reference step satisfies its scalar event $\mathcal{A}_w$. The root events are defined below. The radius satisfies $R_v = R_0 F^{2|v|}$, with R_0 fixed at the root as specified below. For the two calls in this step, put

$$m_{v,1} = m_v, \qquad m_{v,2} = m'_v,$$
$$R_{v,1} = R_v, \qquad R_{v,2} = FR_v, \qquad r_{v,i} = FR_{v,i}/2 \quad (i = 1, 2).$$

Here m'_v is the length of the larger trimmed interval from the first call. Denote the two maximizing ranks, scores, and smaller counts by $k_{v,i}$, $z_{v,i}$, and $q_{v,i} = \min\{k_{v,i}, m_{v,i} - k_{v,i}\}$, respectively. Let $X^{\mathtt{L}}_{v,i}$ and $X^{\mathtt{R}}_{v,i}$ be the two original child bridges in call i, each centered at its own mean and read away from the cutting endpoint. When $q_{v,i} \geq 2r_{v,i}$, define

$$\mathcal{P}_{v,i} = \left\{ \max_{\nu \in \{\mathtt{L},\mathtt{R}\}} \max_{0 \leq j \leq r_{v,i}} \left| X^\nu_{v,i}(j) \right| \leq H_0 \sqrt{r_{v,i}} \right\}.$$

Set $\mathcal{P}_{v,i} = \varnothing$ when $q_{v,i} < 2r_{v,i}$. In the size ranges retained below, this case is excluded for large N. The scalar stage event is

$$\begin{aligned} \mathcal{A}_v = \mathcal{P}_{v,1} \cap \mathcal{P}_{v,2} \\ \cap \left\{ m_v^a \leq q_{v,1} \leq m_v^\beta, \quad v_{m_v}(u_1) < z_{v,1}^2 \leq A^2 L_{m_v} \right\} \\ \cap \left\{ (m'_v)^{2a_2} \leq q_{v,2} \leq (m'_v)^\beta, \quad v_{m'_v}(u_2) < z_{v,2}^2 \leq A^2 L_{m'_v} \right\}. \end{aligned}$$

If the first call fails a size, score, or padding requirement, stop the branch and set $\mathcal{A}_v = \varnothing$. The second-call quantities are accordingly needed only on the complementary event. Let $\mathcal{H}_v^+$ extend $\mathcal{H}_v$ by revealing the two maximizing ranks, their signed heights, and all increments in the discarded middle intervals. Then $\mathcal{A}_v \in \mathcal{H}_v^+$. On $\mathcal{A}_v$, the two retained frontier intervals have conditionally independent interiors given $\mathcal{H}_v^+$, and their endpoint sums, inherited walls, and boundary increments are measurable with respect to this sigma-field. Set $\mathcal{H}_{v0} = \mathcal{H}_{v1} = \mathcal{H}_v^+$ on $\mathcal{A}_v$.

To define the localization event, write $X_{v,i}$ for the centered parent bridge of call i, and let $\Delta^\dagger_{v,i}$ be the $\mathcal{H}_v^+$-measurable bound in (SA-54), evaluated at $m_{v,i}, R_{v,i}, q_{v,i}, z_{v,i}$. Define

$$\mathcal{B}_{v,i} = \left\{ \begin{array}{l} \exists j \in \{1, \dots, m_{v,i} - 1\} : \\ \quad \min\{j, m_{v,i} - j, |j - k_{v,i}|\} > r_{v,i}, \\ \quad z_{v,i} A_{m_{v,i}}(j) - |X_{v,i}(j)| \leq \Delta^\dagger_{v,i} \left(1 + \dfrac{A_{m_{v,i}}(j)}{A_{m_{v,i}}(k_{v,i})} \right) \end{array} \right\},$$

$$\mathcal{R}_v = \mathcal{B}^c_{v,1} \cap \mathcal{B}^c_{v,2}.$$

These events are needed only on $\mathcal{A}_v$. Set $\mathcal{R}_v = \varnothing$ off $\mathcal{A}_v$. We do not add these events or $\mathcal{K}_N$ to $\mathcal{H}_v^+$. The estimates above give, for every admissible $\mathcal{H}_v$,

$$\mathbb{P}\left((\mathcal{A}_v \cap \mathcal{R}_v)^c \,|\, \mathcal{H}_v \right) \leq \varrho.$$

Construct the reference tree using only the Gaussian cuts and padding revelations. Stop each branch when its interval has length at most m_N, the cutoff specified below, or when its scalar event fails. Set $\mathcal{F}_v = \varnothing$ for intervals below the cutoff and define, at every other interval,

$$\mathcal{F}_v = (\mathcal{A}_v \cap \mathcal{R}_v)^c \cup \left(\mathcal{A}_v \cap \mathcal{R}_v \cap \mathcal{F}_{v0} \cap \mathcal{F}_{v1} \right). \qquad \text{[SA-60]}$$

Outside $\mathcal{A}_v$, the child events may be assigned arbitrary values without affecting the definition. Thus $\mathcal{F}_v$ is the event that no path to the cutoff satisfies all scalar and localization requirements. The deterministic bound below ensures that the recursion is finite.

Let I_{v0}, I_{v1} be the retained rank intervals on $\mathcal{A}_v$. Conditional on $\mathcal{H}_v^+$, both intervals and all their boundary data are fixed, and each subsequent operation within child vi uses only these data and $(G_j)_{j\in I_{vi}}$. Finite induction in (SA-60) therefore gives, on $\mathcal{A}_v$,

$$\mathcal{F}_{vi} \in \mathcal{H}_v^+ \vee \sigma(G_j : j \in I_{vi}), \qquad i = 0, 1.$$

The random index set in this display is evaluated at its $\mathcal{H}_v^+$-measurable value. Lemma SA-43 then yields

$$\mathbb{1}(\mathcal{A}_v)\,\mathbb{P}\left(\mathcal{F}_{v0} \cap \mathcal{F}_{v1} \,\middle|\, \mathcal{H}_v^+\right) = \mathbb{1}(\mathcal{A}_v)\,\mathbb{P}\left(\mathcal{F}_{v0} \,\middle|\, \mathcal{H}_v^+\right)\mathbb{P}\left(\mathcal{F}_{v1} \,\middle|\, \mathcal{H}_v^+\right).$$

Suppose the two child failure probabilities are bounded by f_0 uniformly over their admissible histories. Dropping $\mathcal{R}_v$ from the second term of (SA-60) gives

$$\mathbb{P}\left(\mathcal{F}_v \,|\, \mathcal{H}_v\right) \le \mathbb{P}\left((\mathcal{A}_v \cap \mathcal{R}_v)^c \,|\, \mathcal{H}_v\right) + \mathbb{E}\left[\mathbb{1}(\mathcal{A}_v)\,\mathbb{P}\left(\mathcal{F}_{v0} \,\middle|\, \mathcal{H}_v^+\right)\mathbb{P}\left(\mathcal{F}_{v1} \,\middle|\, \mathcal{H}_v^+\right) \,\middle|\, \mathcal{H}_v\right] \le \varrho + f_0^2. \tag{SA-61}$$

The argument does not condition on $\mathcal{R}_v$ or $\mathcal{K}_N$.

The smaller solution of $x = \varrho + x^2$ is

$$f_* = \frac{1 - \sqrt{1 - 4\varrho}}{2} < \frac{1}{2}.$$

At the cutoff the failure probability is zero, and

$$0 \le f_*, \qquad \varrho + f_*^2 = f_*.$$

Backward induction in (SA-61) therefore gives $\mathbb{P}(\mathcal{F}_v \mid \mathcal{H}_v) \le f_*$ at every admissible history. The path-length bound below makes the induction finite. No union bound over reference nodes is used.

The root and the remaining constants. Choose fixed $0 < a_0 < b_0 < 1$. Theorem SA-1, applied to standard normal increments with $p = 1$, gives a limiting probability $b_0 - a_0 > 0$ that the smaller count q is in $[N^{a_0}, N^{b_0}]$. Equations (SA-25) and (SA-26) also give the centered maximum normalization $2L_N + \log L_N + O_{\mathbb{P}}(1)$. Therefore, for a sufficiently large fixed C_{rt}, the event

$$\mathcal{A}_{\mathrm{rt}} = \left\{N^{a_0} \le q \le N^{b_0}, \qquad z^2 \ge 2L_N + \log L_N - C_{\mathrm{rt}}, \qquad z \le A\sqrt{L_N}\right\} \tag{SA-62}$$

has limiting probability bounded below by a fixed positive constant. Indeed,

$$\mathbb{P}(\mathcal{A}_{\mathrm{rt}}) \ge \mathbb{P}(N^{a_0} \le q \le N^{b_0}) - \mathbb{P}(z^2 < 2L_N + \log L_N - C_{\mathrm{rt}}) - \mathbb{P}(z > A\sqrt{L_N}).$$

The first term tends to $b_0 - a_0$, the second is arbitrarily small for large fixed C_{rt}, and the last term is $o(1)$.

At the root there is no padding, and the numerator error is at most $4\Delta_N$. On (SA-62), the exponential prefactor in (SA-55) is at most $C(\log N)^{C'_{\mathrm{rt}}}$ for a fixed $C'_{\mathrm{rt}} > 0$. Choose

$$d_0 > 2C'_{\mathrm{rt}} + 4, \qquad r_{\mathrm{rt}} = \lceil (\log N)^{d_0} \rceil, \qquad R_0 = 2r_{\mathrm{rt}}.$$

The root localization error is therefore at most

$$C(\log N)^{C'_{\mathrm{rt}}} \left\{ \frac{1 + C\log N}{(\log N)^{d_0/2}} + \frac{C(\log N)^2}{N^{a_0/2}} \right\} = o(1).$$

Reveal the middle interval of radius r_{rt} and retain its smaller trimmed child. For the two original centered child bridges, read away from the root cut and denoted by $X^{\mathrm{L}}_{\mathrm{rt}}, X^{\mathrm{R}}_{\mathrm{rt}}$, put, when $q \ge 2r_{\mathrm{rt}}$,

$$\mathcal{P}_{\mathrm{rt}} = \left\{ \max_{\nu\in\{\mathrm{L},\mathrm{R}\}} \max_{0\le j\le r_{\mathrm{rt}}} \left|X^{\nu}_{\mathrm{rt}}(j)\right| \le H_0\sqrt{r_{\mathrm{rt}}} \right\}.$$

Set $\mathcal{P}_{\mathrm{rt}} = \varnothing$ otherwise. The two child-prefix estimates give $\mathbb{P}(\mathcal{P}_{\mathrm{rt}} \mid T_N, k, X_{(0,N]}(k)) \ge 1 - \eta$ on $\mathcal{A}_{\mathrm{rt}}$, by the same choice of H_0. Here $r_{\mathrm{rt}} \le q/2$ for large N on that event. Equations (SA-56)–(SA-58) thus give an initial frontier state with radius R_0. If root localization instead gives an endpoint split, the actual tree has already produced a node of polylogarithmic size.

Fix $C_\beta > 1/|\log\beta|$ and put $D_N = \lceil C_\beta \log\log N \rceil + 2$. Along steps satisfying the scalar requirements, with $m^{(j)}$ the interval length after j steps,

$$\log m^{(j)} \le \beta^j \log N, \qquad \beta^{D_N} \log N \le (\log N)^{1 - C_\beta |\log\beta|} \to 0.$$

Thus every such path crosses any cutoff tending to infinity within D_N steps. The constant C_β, in particular, does not depend on the cutoff exponent. All radii used on these paths are, in turn, bounded by

$$R_* = R_0 F^{2D_N+2} \le (\log N)^{r_*}$$

for a fixed r_*, increasing it once to absorb multiplicative constants. With these constants fixed, choose the exponent B so large that

$$Ba_2 > r_* + 4, \qquad B > r_* + 2, \qquad m_N = (\log N)^B.$$

We impose the estimates for intermediate sizes at least $m_N/2$ to cover the second call of a step beginning just above m_N. Its size may be slightly smaller than the cutoff, but the factor of two does not change any power bound.

For $m \geq m_N/2$ and $q \geq cm^{a_2}$,

$$\frac{RL_m}{m^{a_2}} \leq C(\log N)^{r_* - Ba_2} \log\log N = o(1), \qquad \frac{r}{q} = o(1), \qquad z\sqrt{R/q} = o(1).$$

These bounds control the trimming, slope, and mean corrections. For either tested exponent $\vartheta \geq a_2$, the intercept satisfies

$$\frac{\delta\sqrt{L_m}}{\sqrt{\lceil m^\vartheta \rceil}} \leq C\sqrt{L_m}\, m^{-(a_2 - r_*/B)/2} \to 0.$$

Since the right side is eventually decreasing, it gives an envelope ω_m satisfying (SA-51) uniformly over the admissible histories. The same choice of cutoff gives

$$\frac{C\sqrt{R_*}(1+\log N)}{\sqrt{m_N^{a_2}}} \leq C(\log N)^{r_*/2 + 1 - Ba_2/2} = o(1),$$

which bounds the second term of (SA-55) uniformly. Thus (SA-61) applies to the entire finite reference tree above the cutoff. The constants were chosen in order, beginning with the two-split exponents and thresholds, then H_0, P, F, the root radius exponent, the radius-growth bound, and finally B. All earlier choices are therefore independent of both B and N.

Let $\mathcal{F}_{\mathrm{rt,sm}}$ denote the failure event (SA-60) for the retained smaller root interval. Conditional on the root rank, signed height, total sum, and revealed middle increments, its failure probability is at most f_* on $\mathcal{A}_{\mathrm{rt}} \cap \mathcal{P}_{\mathrm{rt}}$. Therefore

$$\mathbb{P}\big(\mathcal{A}_{\mathrm{rt}} \cap \mathcal{P}_{\mathrm{rt}} \cap \mathcal{F}_{\mathrm{rt,sm}}^c\big) \geq (1 - f_*)(1 - \eta)\mathbb{P}(\mathcal{A}_{\mathrm{rt}}).$$

Intersect with root localization and $\mathcal{K}_N$, losing $o(1)$ in probability. A successful matched path then produces either an endpoint child or an enlarged terminal interval. Its actual size and depth obey

$$n_{\mathrm{node}} \leq \max\{CR_*, m_N + 2R_*\} \leq 2m_N, \qquad d_{\mathrm{node}} \leq 1 + 2D_N \leq \lceil C_0 \log\log N \rceil$$

for any fixed $C_0 > 2C_\beta$ and all large N. We may therefore take

$$q_{\mathrm{node}} = \tfrac{1}{2}(1 - f_*)(1 - \eta) \liminf_{N\to\infty} \mathbb{P}(\mathcal{A}_{\mathrm{rt}}) > 0.$$

The construction uses only the ordered residuals and the auxiliary Gaussian sequence, giving the stated measurability. □

Transfer to the honest estimator. To prove Theorem SA-7, apply Lemma SA-47. There are consequently fixed constants $B, C_0, q_{\mathrm{node}} > 0$ such that, for all sufficiently large N, with probability at least q_{node}, the actual construction tree contains a node of size at most $2(\log N)^B$ at depth at most $\lceil C_0 \log\log N \rceil$. If n_j is the count after j successive smaller-child choices, with left ties, then

$$n_{j+1} \leq \lfloor n_j/2 \rfloor, \qquad n_j \leq 2^{-j}\, 2(\log N)^B.$$

Thus a singleton is reached within $\lceil \log_2\{2(\log N)^B\} \rceil$ further splits. Put

$$k_N^\star = \lceil C_0 \log\log N \rceil + \lceil \log_2\{2(\log N)^B\} \rceil.$$

Write $U_{(1)} < \cdots < U_{(N)}$ for the ordered construction covariates after the probability integral transform, and put $U_{(0)} = 0$ and $U_{(N+1)} = 1$. Let $\mathcal{P}_N^\star$ be the terminal partition of the actual maximal tree truncated at depth $k_N^\star$, and set

$$\mathcal{E}_N = \left\{ \min_{\mathsf{t} \in \mathcal{P}_N^\star} \sum_{i=1}^N \mathbb{1}(U_{(i)} \in \mathsf{t}) = 1 \right\}.$$

The deterministic descent just proved gives

$$\liminf_{N\to\infty} \mathbb{P}(\mathcal{E}_N) \geq q_{\mathrm{node}} > 0. \qquad \text{[SA-63]}$$

For any fixed $C_1 > C_0 + B/\log 2$, every depth sequence in the theorem satisfies $K_N \geq k_N^\star$ eventually. The capped and uncapped trees therefore have the same nodes through depth $k_N^\star$. A construction singleton, moreover, has no valid split and remains a terminal node at all larger maximal depths.

On $\mathcal{E}_N$, select the singleton with the smallest global construction rank, given by

$$J_N = \min\left\{j \in [N] : \exists \mathsf{t} \in \mathcal{P}_N^\star, \quad \{i : U_{(i)} \in \mathsf{t}\} = \{j\}\right\}.$$

Set $J_N = 1$ on $\mathcal{E}_N^c$. The scalar CART scores and prescribed tie rule use only residuals and cut ranks. Hence

$$\mathcal{E}_N \in \sigma(Z_1, \ldots, Z_N), \qquad J_N \text{ is } \sigma(Z_1, \ldots, Z_N)\text{-measurable.}$$

Consequently

$$(\mathbb{1}(\mathcal{E}_N), J_N) \perp\!\!\!\perp (U_{(1)}, \ldots, U_{(N)}). \tag{SA-64}$$

Both $\mathcal{E}_N$ and J_N are defined from the actual tree, so this selection does not use the auxiliary Gaussian variables.

Under the observed-value threshold convention, a singleton of rank $j < N$ is the cell $(U_{(j-1)}, U_{(j)}]$, with its left endpoint included when $j = 1$. The singleton of rank N, by contrast, is $(U_{(N-1)}, 1]$. Indeed any boundary separating two adjacent construction ranks is the observation with the smaller rank, and the rightmost cell extends to the boundary of the covariate support. Define

$$V_N = \begin{cases} U_{(J_N)} - U_{(J_N-1)}, & J_N < N, \\ 1 - U_{(N-1)}, & J_N = N. \end{cases}$$

On $\mathcal{E}_N$, this is the population mass of the selected leaf. Off $\mathcal{E}_N$, it defines the corresponding rank interval. The joint density of the N uniform order statistics is $N!$ on their ordered simplex. Changing variables to the spacings then has Jacobian one and gives the uniform law on the unit simplex [16, Chapter V, Theorem 2.1, pp. 207–208]. Integrating that density gives

$$f_{U_{(j)}-U_{(j-1)}}(v) = N(1-v)^{N-1}, \qquad j \le N,$$
$$f_{1-U_{(N-1)}}(v) = N(N-1)v(1-v)^{N-2}, \qquad 0 < v < 1.$$

Thus, conditional on $\mathcal{E}_N$ and $J_N = j$,

$$V_N \sim \begin{cases} \mathsf{Beta}(1, N), & j < N, \\ \mathsf{Beta}(2, N-1), & j = N. \end{cases} \tag{SA-65}$$

The second case combines the last two spacings. The conditional laws follow from (SA-64), without conditioning on the numerical construction covariates.

Let Q_N be the number of estimation observations in this rank interval. Conditional on the construction data, $Q_N \sim \mathsf{Binomial}(M, V_N)$. Integrating over (SA-65) gives the exact formulas

$$\mathbb{P}\left(Q_N = 1 \,|\, \mathcal{E}_N,\, J_N = j\right) = M\mathbb{E}\left[V_N(1-V_N)^{M-1} \,\middle|\, \mathcal{E}_N,\, J_N = j\right] = \begin{cases} \dfrac{MN}{(M+N)(M+N-1)}, & j < N, \\ \dfrac{2MN(N-1)}{(M+N)(M+N-1)(M+N-2)}, & j = N. \end{cases} \tag{SA-66}$$

For $N \ge 2$, the two expressions are bounded below, respectively, by

$$\frac{M/N}{(1+M/N)^2}, \qquad \frac{2(M/N)(1-1/N)}{(1+M/N)^3}.$$

Since $\rho \le M/N \le \rho^{-1}$ and $N \ge 2$,

$$\min\left\{\frac{M/N}{(1+M/N)^2}, \frac{2(M/N)(1-1/N)}{(1+M/N)^3}\right\} \ge q_\rho := \frac{\rho}{(1+\rho^{-1})^3} > 0.$$

No convergence of M/N is therefore required.

Choose $\epsilon > 0$ such that $q_\epsilon = \mathbb{P}(|\varepsilon_i| > \epsilon) > 0$, which is possible because $\mathbb{V}[\varepsilon_i] > 0$. Conditional on all construction data and all estimation covariates, the estimation residuals remain independent with their original law. On $\mathcal{E}_N \cap \{Q_N = 1\}$, the selected terminal estimate is therefore $\mu + \varepsilon_i$ for its unique estimation observation. Let I_N be that estimation index when $Q_N = 1$, and put $I_N = 1$ otherwise. If $\mathscr{G}$ is the sigma-field generated by the construction data and estimation covariates, then I_N and $\mathcal{E}_N \cap \{Q_N = 1\}$ are $\mathscr{G}$-measurable, and

$$\mathbb{E}\left[\mathbb{1}(\mathcal{E}_N)\mathbb{1}(Q_N = 1)\mathbb{1}(|\varepsilon_{I_N}| > \epsilon) \,|\, \mathscr{G}\right] = q_\epsilon \mathbb{1}(\mathcal{E}_N)\mathbb{1}(Q_N = 1).$$

Here ε_{I_N} is an estimation-fold residual. It follows from (SA-63) and (SA-66) that

$$\liminf_{N\to\infty} \mathbb{P}\left(\sup_{\mathbf{x}\in\mathcal{X}} \left|\hat{\mu}^{\mathtt{HON}}(\mathbf{x}) - \mu(\mathbf{x})\right| > \epsilon\right) \ge q_{\text{node}} q_\rho q_\epsilon > 0.$$

Because the selected leaf contains an estimation observation, the conclusion holds under any convention for empty estimation leaves. The construction singleton cannot be split, so the inconsistency conclusion persists at every larger maximal depth. For the particular choice $K_N = \lceil C_1 \log\log N \rceil$, $2^{\lceil C_1 \log\log N\rceil}/N \le 2(\log N)^{C_1 \log 2}/N \to 0$. Since the fold sizes are comparable, Theorem SA-6 then gives integrated L^q consistency for every fixed $1 \le q < \infty$.

SA-5.9. Proof of Lemma SA-8. For the honest estimator, condition on the selected tree partition and on the covariates in the final estimation sample. Write $\tilde{\mathbf{X}}$ for its covariate matrix. If the terminal node containing $\mathbf{x}$ has positive final estimation count, the conditional expectation of the terminal average is μ. If the count is zero, the estimator is defined to be zero. Thus, with $\mathscr{G} = \sigma(\mathcal{D}_{\mathsf{T}}, \tilde{\mathbf{X}})$,

$$\mathbb{E}[\hat{\mu}^{\texttt{HON}}(\mathbf{x}) \mid \mathscr{G}] = \mu(\mathbf{x})\mathbb{1}(N_{\texttt{HON}}(\mathbf{x}) > 0).$$

Taking expectations then gives the displayed empty cell expression.

For NSS, write $\boldsymbol{\varepsilon} = (\varepsilon_1, \ldots, \varepsilon_n)^\top$. Conditional on the covariates,

$$\boldsymbol{\varepsilon} \overset{d}{=} -\boldsymbol{\varepsilon}.$$

By the assumed partition invariance,

$$\mathcal{P}(\boldsymbol{\varepsilon}) = \mathcal{P}(-\boldsymbol{\varepsilon}), \qquad \hat{\mu}^{\texttt{NSS}}(\mathbf{x}; -\boldsymbol{\varepsilon}) - \mu(\mathbf{x}) = -\{\hat{\mu}^{\texttt{NSS}}(\mathbf{x}; \boldsymbol{\varepsilon}) - \mu(\mathbf{x})\}.$$

Every NSS leaf is also nonempty. Moreover,

$$|\hat{\mu}^{\texttt{NSS}}(\mathbf{x}) - \mu(\mathbf{x})| \leq \max_{1\leq i\leq n} |\varepsilon_i| \leq \sum_{i=1}^{n} |\varepsilon_i|, \qquad \sum_{i=1}^{n} \mathbb{E}[|\varepsilon_i|] < \infty.$$

Conditional symmetry and integrability together give

$$\mathbb{E}[\hat{\mu}^{\texttt{NSS}}(\mathbf{x}) - \mu(\mathbf{x}) \mid \mathbf{X}] = -\mathbb{E}[\hat{\mu}^{\texttt{NSS}}(\mathbf{x}) - \mu(\mathbf{x}) \mid \mathbf{X}] = 0.$$

Iterated expectations then prove the assertion.

More generally, let $\mathscr{G}$ be a sigma field and let a residual vector $\boldsymbol{e}$ have a conditionally symmetric law, $\mathcal{L}(\boldsymbol{e} \mid \mathscr{G}) = \mathcal{L}(-\boldsymbol{e} \mid \mathscr{G})$. Suppose an estimator has the form $\theta H + W$, where θ is constant, $H \in \{0, 1\}$ is unchanged by $\boldsymbol{e} \mapsto -\boldsymbol{e}$, and W is integrable and changes sign under this map, with all $\mathscr{G}$-measurable inputs held fixed. Then

$$\mathbb{E}[W \mid \mathscr{G}] = -\mathbb{E}[W \mid \mathscr{G}] = 0, \qquad \mathbb{E}[\theta H + W] = \theta\mathbb{P}(H = 1).$$

This conditional sign argument also handles the causal estimators below.

SA-5.10. Proof of Corollary SA-9. Let $\tilde{y}_i = y_i(d_i - \xi)/\{\xi(1-\xi)\}$ and $\tilde{\varepsilon}_i = \tilde{y}_i - \tau$. Under Assumption SA-2, $\mathbb{E}[\tilde{y}_i \mid \mathbf{x}_i] = \tau$, $\tilde{\varepsilon}_i$ is independent of $\mathbf{x}_i$, and $\tilde{\varepsilon}_i$ satisfies the same subexponential and nondegeneracy conditions needed in Assumption SA-1. Equations (SA-8)–(SA-10) verify these moment and independence conditions. For every candidate child,

$$\hat{\tau}_{\texttt{IPW}}(\mathsf{t}) = \frac{1}{n(\mathsf{t})} \sum_{\mathbf{x}_i \in \mathsf{t}} \tilde{y}_i,$$

so substitution in (SA-6) gives the regression CART impurity gain and valid candidate set, with the same coordinate/cut-rank tie rule. Thus NSS-IPW is CART applied to $\tilde{y}_i$, and Theorem SA-1 gives the stated split index result.

SA-5.11. Proof of Corollary SA-10. With the same transformed outcome $\tilde{y}_i = y_i(d_i - \xi)/\{\xi(1-\xi)\}$, the IPW terminal node estimate in a stump is the corresponding left or right node average of $\tilde{y}_i$. Hence Theorem SA-2 applied to $\tilde{y}_i$ gives uniform consistency, the exact uniform-error probability limit, and the boundary lower bound, with $\sigma^2 = \mathbb{V}[\tilde{\varepsilon}_i]$. The boundary statement follows after the same marginal probability integral transform normalization used in Theorem SA-2.

SA-5.12. Proof of Corollary SA-11. Apply Theorem SA-3 to the transformed outcomes $\tilde{y}_i = y_i(d_i - \xi)/\{\xi(1-\xi)\}$. Their mean is τ, and their centered residuals satisfy Assumption SA-1. Every IPW terminal estimate is exactly the corresponding transformed outcome average, so the regression theorem gives the result, since $\sqrt{2 + o(1)} > 1$ eventually.

SA-5.13. Proof of Corollary SA-12. Let $\tilde{y}_i = y_i(d_i - \xi)/\{\xi(1-\xi)\}$ and $\tilde{\varepsilon}_i = \tilde{y}_i - \tau$. Under Assumption SA-2, the transformed residuals are centered, independent of the covariates, and subexponential. On a nonempty leaf, the NSS-IPW estimate is the sample average of $\tilde{y}_i$. Since the proof of Theorem SA-4 bounds all rectangle averages simultaneously, it applies to any selected partition. Use the residuals $\tilde{\varepsilon}_i$ in that argument to obtain both bounds, with the stated dependence of the constants, for every fixed $1 \leq q < \infty$.

SA-5.14. Proof of Corollary SA-13. Conditional on the construction fold, the honest IPW estimator averages $\tilde{y}_i$ over the independent estimation fold in the selected terminal node. Thus Theorem SA-5 applies to the transformed outcome regression problem.

SA-5.15. Proof of Corollary SA-14. Let $\tilde{y}_i = y_i(d_i - \xi)/\{\xi(1-\xi)\} = \tau + \tilde{\varepsilon}_i$. Under Assumption SA-2, these transformed outcomes have mean τ, variance $\mathbb{V}[\tilde{\varepsilon}_i]$, and are independent of the covariates. Conditional on the construction fold, the terminal partition $\mathcal{P}$ is independent of the transformed outcomes and covariates in the estimation fold. For $q = 2$, the scalar calculation in the proof of Theorem SA-6 therefore gives

$$\mathbb{E}\left[\int_{\mathcal{X}} (\check{\tau}_{\mathtt{IPW}}(\mathbf{x}) - \tau(\mathbf{x}))^2 \, \mathrm{d}F_X(\mathbf{x}) \,\middle|\, \mathcal{D}_{\mathtt{T}}\right] \leq \frac{2\mathbb{V}[\tilde{\varepsilon}_i] + \tau^2}{n_\tau + 1} |\mathcal{P}|.$$

The τ^2 term accounts for the zero estimate in an empty estimation leaf. Unconditioning, using $|\mathcal{P}| \leq 2^K$, and taking the square root proves the $q = 2$ moment bound with $C_1 = (2\mathbb{V}[\tilde{\varepsilon}_i] + \tau^2)^{1/2}$. Applying Markov's inequality to integrated squared error and taking square roots then gives the probability bound with $C_2 = C_1$. For the remaining values of q, apply the same scalar argument to the transformed residuals. Their exponential moment condition gives the required moments for $q > 2$, including the empty-leaf term $|\tau|^q$. Hölder's inequality and the $q = 2$ bounds together give the results for $1 \leq q < 2$.

SA-5.16. Proof of Corollary SA-15. Apply Theorem SA-7 to the transformed outcomes $\tilde{y}_i = y_i(d_i - \xi)/\{\xi(1-\xi)\} = \tau + \tilde{\varepsilon}_i$. Section SA-3.1 verifies Assumption SA-1 for this regression model. The IPW split criterion, prescribed tie rule, and terminal estimates agree with regression CART applied to these outcomes, and the maximal-growth and fold-balance conditions coincide.

SA-5.17. Proof of Lemma SA-16. Fix one coordinate ℓ, suppress its index, and order the data as $x_1 \leq \cdots \leq x_n$. A union bound over the fixed p coordinates will then give the result. The covariates are independent of the treatment assignments and potential errors, so ordering by a covariate preserves their joint i.i.d. law. We may assume the retained range is nonempty and evaluate every expression with an empirical arm denominator only at valid candidates.

For $d \in \{0, 1\}$, write

$$A_{d,k} = \sum_{i=1}^{k} \mathbb{1}(d_i = d)\varepsilon_i(d), \qquad N_{d,k} = \sum_{i=1}^{k} \mathbb{1}(d_i = d), \qquad B_{d,k} = N_{d,k} - \xi_d k.$$

On a valid left child,

$$\hat{\tau}_{\mathtt{L}}^{\mathtt{DIM}}(k) - \tau = \frac{A_{1,k}}{N_{1,k}} - \frac{A_{0,k}}{N_{0,k}}, \qquad \bar{\tau}_{\mathtt{L}}^{\mathtt{IPW}}(k) = \frac{A_{1,k}}{\xi_1 k} - \frac{A_{0,k}}{\xi_0 k}.$$

The corresponding right-child expressions use $A_{d,n} - A_{d,k}$ and $N_{d,n} - N_{d,k}$, with k replaced by $n - k$. In particular, the two appearances of the constant treatment effect cancel from the left–right DIM contrast.

Define the denominator event

$$\mathcal{A}_n(r) = \bigcap_{d=0}^{1} \bigcap_{r \leq k \leq n-r} \left\{ N_{d,k} \geq \frac{\xi_d k}{2}, \quad N_{d,n} - N_{d,k} \geq \frac{\xi_d (n-k)}{2} \right\}, \tag{SA-67}$$

where an intersection over an empty integer range is the whole sample space. Lemma SA-40, applied to the forward and reversed orderings in both arms, gives $\mathbb{P}(\mathcal{A}_n(r_n)^c) = O(r_n^{-1})$. Every retained candidate is, moreover, valid on $\mathcal{A}_n(r_n)$. For a left-child term on this event,

$$\sqrt{k}\left|\frac{A_{d,k}}{N_{d,k}} - \frac{A_{d,k}}{\xi_d k}\right| = \frac{|A_{d,k} B_{d,k}|}{\xi_d \sqrt{k} N_{d,k}} \leq \frac{2}{\xi_d^2} \frac{|A_{d,k} B_{d,k}|}{k^{3/2}}.$$

Its maximum over $k \geq r_n$ is $O_{\mathbb{P}}(r_n^{-1/2})$ by the first product bound in that lemma. Reversing the order similarly proves the same bound for each right-child term, with weight $\sqrt{n-k}$.

Expand the square to compare the two splitting criteria. Define

$$R_1(k) = \hat{\tau}_{\mathtt{L}}^{\mathtt{DIM}}(k) - \hat{\tau}_{\mathtt{R}}^{\mathtt{DIM}}(k) + \bar{\tau}_{\mathtt{L}}^{\mathtt{IPW}}(k) - \bar{\tau}_{\mathtt{R}}^{\mathtt{IPW}}(k),$$
$$R_2(k) = \hat{\tau}_{\mathtt{L}}^{\mathtt{DIM}}(k) - \hat{\tau}_{\mathtt{R}}^{\mathtt{DIM}}(k) - \bar{\tau}_{\mathtt{L}}^{\mathtt{IPW}}(k) + \bar{\tau}_{\mathtt{R}}^{\mathtt{IPW}}(k).$$

Since

$$\sqrt{\frac{k(n-k)}{n}} \leq \min\{\sqrt{k}, \sqrt{n-k}\},$$

the four preceding arm/child bounds imply

$$\max_{\substack{r_n \leq k \leq n - r_n \\ (k,\ell) \in \mathcal{V}_{\mathtt{DIM}}}} \sqrt{\frac{k(n-k)}{n}} |R_2(k)| = O_{\mathbb{P}}(r_n^{-1/2}).$$

Put

$$Z_i = \frac{d_i \varepsilon_i(1)}{\xi} - \frac{(1-d_i)\varepsilon_i(0)}{1-\xi}, \qquad S_k = \sum_{i=1}^{k} Z_i.$$

These increments are centered and i.i.d., with a finite exponential moment near zero and

$$\mathbb{V}[Z_i] = \frac{\mathbb{V}[\varepsilon_i(1)]}{\xi} + \frac{\mathbb{V}[\varepsilon_i(0)]}{1-\xi} > 0,$$

since $d_i(1-d_i) = 0$. Equation (SA-23), applied in both orders, then gives

$$\max_{1\le k<n} \sqrt{\frac{k(n-k)}{n}} |\bar{\tau}_{\mathtt{L}}^{\mathtt{IPW}}(k) - \bar{\tau}_{\mathtt{R}}^{\mathtt{IPW}}(k)| \le \max_{1\le k<n} \frac{|S_k|}{\sqrt{k}} + \max_{1\le k<n} \frac{|S_n - S_k|}{\sqrt{n-k}} = O_{\mathbb{P}}(\sqrt{\log\log n}).$$

For $R_1(k)$, the definitions give the identity

$$R_1(k) = R_2(k) + 2\big(\bar{\tau}_{\mathtt{L}}^{\mathtt{IPW}}(k) - \bar{\tau}_{\mathtt{R}}^{\mathtt{IPW}}(k)\big).$$

Consequently,

$$\max_{\substack{r_n \le k \le n-r_n \\ (k,\ell)\in\mathcal{V}_{\mathtt{DIM}}}} \sqrt{\frac{k(n-k)}{n}} |R_1(k)| = O_{\mathbb{P}}(\sqrt{\log\log n}).$$

For each valid candidate k,

$$\mathscr{I}^{\mathtt{DIM}}(k) - \bar{\mathscr{I}}^{\mathtt{IPW}}(k) = \frac{k(n-k)}{n} R_1(k) R_2(k).$$

The product of the two weighted bounds is therefore $O_{\mathbb{P}}(\sqrt{\log\log n / r_n})$, as required. All retained candidates are, as noted above, valid on the denominator event. Since the denominator event has probability tending to one, the same stochastic order holds for the maximum over valid candidates. A finite union over coordinates finally proves the result.

SA-5.18. Proof of Lemma SA-17. Fix one coordinate ℓ, suppress its index, and order the data as $x_1 \le \cdots \le x_n$. A union bound over the fixed p coordinates will then give the result. First consider $1 \le k \le s_n$. Retain the arm notation $A_{d,k}, N_{d,k}$ from the preceding proof. Given the assignments, the observed errors in arm d retain their original i.i.d. law in order of appearance. We may therefore represent them by an i.i.d. stream, independent of the assignments, whose partial sums $T_d(h)$ satisfy $A_{d,k} = T_d(N_{d,k})$. Each individual contributes only one observed error, so the representation allows dependence between that individual's potential errors.

Use an auxiliary infinite i.i.d. Bernoulli assignment sequence with probability ξ_d. The strong law of large numbers [18, Theorem 2.4.1] implies

$$C_d := \sup_{\substack{k\ge 1 \\ N_{d,k}>0}} \frac{k}{N_{d,k}} < \infty \quad \text{almost surely.}$$

Indeed, for an almost-surely finite K_d,

$$C_d \le \max\left\{ \frac{2}{\xi_d}, \max_{1\le k\le K_d:\, N_{d,k}>0} \frac{k}{N_{d,k}} \right\} < \infty, \qquad C_d = O_{\mathbb{P}}(1).$$

Equation (SA-23), applied to the arm stream at horizon s_n, then gives

$$\max_{1\le k\le s_n : N_{d,k}>0} \sqrt{k} \left| \frac{A_{d,k}}{N_{d,k}} \right| \le C_d^{1/2} \max_{1\le h\le \lfloor s_n \rfloor} \frac{|T_d(h)|}{\sqrt{h}} = O_{\mathbb{P}}(\sqrt{\log\log s_n}).$$

The bound does not, however, require the two factors to be independent. Because the auxiliary construction has the same finite-sample law as the ordered observations, it gives the bound for every n.

For the larger right child, define

$$\mathcal{A}_n^{\mathtt{R}}(s_n) = \bigcap_{d=0}^{1} \left\{ \min_{1\le k\le s_n} (N_{d,n} - N_{d,k}) \ge \xi_d n/2 \right\}.$$

Since $N_{d,n}/n \to_{\mathbb{P}} \xi_d$ and $N_{d,k} \le s_n = o(n)$, we have $\mathbb{P}(\mathcal{A}_n^{\mathtt{R}}(s_n)) \to 1$. Doob's L^2 maximal inequality (SA-17) also gives

$$\max_{1\le k\le s_n} |A_{d,n} - A_{d,k}| \le |A_{d,n}| + \max_{1\le k\le s_n} |A_{d,k}| = O_{\mathbb{P}}(\sqrt{n}).$$

Together with the denominator bound, this gives

$$\max_{\substack{1\le k\le s_n\\(k,\ell)\in\mathcal{V}_{\mathtt{DIM}}}} \sqrt{\frac{k(n-k)}{n}}\left|\frac{A_{d,k}}{N_{d,k}}\right| = O_{\mathbb{P}}(\sqrt{\log\log s_n}),$$

$$\max_{\substack{1\le k\le s_n\\(k,\ell)\in\mathcal{V}_{\mathtt{DIM}}}} \sqrt{\frac{k(n-k)}{n}}\left|\frac{A_{d,n}-A_{d,k}}{N_{d,n}-N_{d,k}}\right| = O_{\mathbb{P}}(\sqrt{s_n/n}),$$

where empirical denominators are evaluated only at valid candidates. The constant arm means cancel, so $(\sum_{h=1}^4 a_h)^2 \le 4\sum_{h=1}^4 a_h^2$ gives

$$\max_{\substack{1\le k\le s_n\\(k,\ell)\in\mathcal{V}_{\mathtt{DIM}}}} \mathscr{I}^{\mathtt{DIM}}(k) = O_{\mathbb{P}}(\log\log s_n + s_n/n).$$

For the proxy, apply (SA-23) directly to the i.i.d. centered increments $d_i\varepsilon_i(1)/\xi - (1-d_i)\varepsilon_i(0)/(1-\xi)$. The left-child bound then holds without empirical denominators, while the maximal inequality and $n-k \ge n-s_n$ give the same bound for the right child. Thus the displayed bound also holds for the proxy maximum, including candidates that are invalid for DIM. Since $\log\log s_n \to \infty$ and $s_n/n \to 0$, the resulting order is $O_{\mathbb{P}}(\log\log s_n)$. Apply the argument in the reversed sample ordering for $n - s_n \le k < n$ and take a union bound over the fixed p coordinates.

SA-5.19. Proof of Theorem SA-18. Let $r_n = \lceil (\log n)^8 \rceil$ and $\sigma^2 = \mathbb{V}[d_i\varepsilon_i(1)/\xi - (1-d_i)\varepsilon_i(0)/(1-\xi)] > 0$. Lemma SA-16 gives

$$\max_{\substack{1\le\ell\le p,\ r_n\le k\le n-r_n\\(k,\ell)\in\mathcal{V}_{\mathtt{DIM}}}} |\mathscr{I}^{\mathtt{DIM}}(k,\ell) - \bar{\mathscr{I}}^{\mathtt{IPW}}(k,\ell)| = o_{\mathbb{P}}(1).$$

All candidates in this retained range are valid with probability tending to one. The proof of Lemma SA-17, applied with $s_n = r_n$, also bounds each discarded maximum by $O_{\mathbb{P}}(\log\log r_n) = o_{\mathbb{P}}(\log\log n)$.

The centered proxy increments are independent of the covariates, have variance σ^2, and have a finite exponential moment near zero. The one-coordinate approximation in the proof of Theorem SA-1 applies to the standardized proxy increments. Thus Lemma SA-39 applies to their joint block maxima. For a fixed logarithmic subdivision $\mathcal{B}_{1,n},\dots,\mathcal{B}_{J,n}$ of $\{\lceil r_n\rceil,\dots,\lfloor n-r_n\rfloor\}\times[p]$, with $J\ge 2$, each block lying in one coordinate and having positive limiting length under $k\mapsto \log\{k/(n-k)\}/\log n$, put

$$M_{j,n} = \frac{1}{2}\left\{\max_{(k,\ell)\in\mathcal{B}_{j,n}\cap\mathcal{V}_{\mathtt{DIM}}} \frac{\mathscr{I}^{\mathtt{DIM}}(k,\ell)}{\sigma^2} - \beta_n\right\} \quad\text{and}\quad \widetilde{M}_{j,n} = \frac{1}{2}\left\{\max_{(k,\ell)\in\mathcal{B}_{j,n}\cap\mathcal{V}_{\mathtt{DIM}}} \frac{\bar{\mathscr{I}}^{\mathtt{IPW}}(k,\ell)}{\sigma^2} - \beta_n\right\},$$

where β_n is defined in (SA-25). Assign both quantities zero if the indexing set is empty. Such empty sets have, however, probability $o(1)$. The score approximation gives

$$E_n := \max_{1\le j\le J} |M_{j,n} - \widetilde{M}_{j,n}| = o_{\mathbb{P}}(1), \qquad (\widetilde{M}_{1,n},\dots,\widetilde{M}_{J,n}) \rightsquigarrow (Y_1,\dots,Y_J),$$

with independent continuous limiting variables. Choose $J_n \in \arg\max_{1\le j\le J} M_{j,n}$ and $\widetilde{J}_n \in \arg\max_{1\le j\le J}\widetilde{M}_{j,n}$. Let D_n be the difference between the largest and second largest proxy block maxima, and let D be the corresponding limiting gap. The gap map is continuous. The continuous-mapping and closed-set consequences of [18, Theorem 3.10.1] then give

$$D_n \rightsquigarrow D, \qquad \mathbb{P}(D=0)=0, \qquad \limsup_{n\to\infty}\mathbb{P}(D_n\le\eta)\le\mathbb{P}(D\le\eta).$$

Here $\mathbb{P}(D=0)=0$ follows from independence and continuity of the limiting coordinates. Since a change of winning block requires $D_n \le 2E_n$,

$$\mathbb{P}(J_n\ne\widetilde{J}_n)\le\mathbb{P}(2E_n>\eta)+\mathbb{P}(D_n\le\eta), \qquad \limsup_{n\to\infty}\mathbb{P}(J_n\ne\widetilde{J}_n)\le\mathbb{P}(D\le\eta)\to 0 \qquad (\eta\downarrow 0).$$

As in Lemma SA-37, ties within a block do not affect this argument. The retained maximum is $(2\sigma^2+o_{\mathbb{P}}(1))\log\log n$, excluding the discarded endpoints with probability tending to one. The target block $[n^a,n^b]$ in coordinate ℓ is therefore selected with limiting probability $(b-a)/(2p)$. Reversing the order similarly proves the right-boundary assertion, and the probability of no valid split is $o(1)$. The conditioning argument in the proof of Theorem SA-1 then gives $\mathbb{P}(\hat{\jmath}_{\mathtt{DIM}}=\ell)\to 1/p$ and the conditional window limits.

SA-5.20. Proof of Theorem SA-19. Work on $\mathcal{V}_{\mathtt{DIM}} \neq \varnothing$, whose probability tends to one. Write $m = \min\{\hat{\imath}_{\mathtt{DIM}}, n - \hat{\imath}_{\mathtt{DIM}}\}$ and use S, S^c for the smaller and larger selected root children, taking the left child as S in a tie. The selected-child expressions below are used only on this event. By the definition of the DIM criterion,

$$\sqrt{m}\,|\hat{\tau}_{\mathtt{DIM}}(\mathtt{t}_S) - \hat{\tau}_{\mathtt{DIM}}(\mathtt{t}_{S^c})| = \sqrt{\frac{n}{n-m}\,\mathscr{I}^{\mathtt{DIM}}(\hat{\imath}_{\mathtt{DIM}}, \hat{\jmath}_{\mathtt{DIM}})}.$$

As in (SA-27), this uses only that the criterion is a squared difference of child CATE estimates. Uniformly over all root cuts whose larger child has at least $n/2$ observations, (SA-17) bounds each arm residual sum by $O_{\mathbb{P}}(\sqrt{n})$. Lemma SA-40, applied with $r = n/2$ in the forward and reversed orders, also bounds the Bernoulli counts in that child below by a fixed positive multiple of n with probability tending to one. Hence

$$|\hat{\tau}_{\mathtt{DIM}}(\mathtt{t}_{S^c}) - \tau| = O_{\mathbb{P}}(n^{-1/2}).$$

The preceding split proof also gives $\max_{(k,\ell)\in\mathcal{V}_{\mathtt{DIM}}} \mathscr{I}^{\mathtt{DIM}}(k,\ell) = (2\sigma^2 + o_{\mathbb{P}}(1)) \log\log n$.

Theorem SA-18, summed over coordinates and both boundaries, gives $\log m / \log n \rightsquigarrow \mathsf{Uniform}(0,1)$, hence $m/n \to_{\mathbb{P}} 0$ and $m/\log\log n \to_{\mathbb{P}} \infty$. The score identity and the preceding bounds therefore give

$$\frac{\sqrt{m}}{\sigma\sqrt{2\log\log n}} \sup_{\mathbf{x}\in\mathcal{X}} |\hat{\tau}_{\mathtt{DIM}}(\mathbf{x}) - \tau(\mathbf{x})| \to_{\mathbb{P}} 1.$$

Uniform consistency and the exact probability limit then follow by the same argument as in the proof of Theorem SA-2.

On $\{m \le n^b\}$, the reverse triangle inequality gives

$$\begin{aligned}\left|\sqrt{m}\,|\hat{\tau}_{\mathtt{DIM}}(\mathtt{t}_S) - \tau| - \sqrt{\mathscr{I}^{\mathtt{DIM}}(\hat{\imath}_{\mathtt{DIM}}, \hat{\jmath}_{\mathtt{DIM}})}\right| &\le \sqrt{m}\,|\hat{\tau}_{\mathtt{DIM}}(\mathtt{t}_{S^c}) - \tau| + \left(\sqrt{\frac{n}{n-m}} - 1\right)\sqrt{\mathscr{I}^{\mathtt{DIM}}(\hat{\imath}_{\mathtt{DIM}}, \hat{\jmath}_{\mathtt{DIM}})} \\ &= O_{\mathbb{P}}(n^{(b-1)/2}) + O_{\mathbb{P}}(n^{b-1}\sqrt{\log\log n}) = o_{\mathbb{P}}(1).\end{aligned}$$

It follows, uniformly on this event, that

$$\sqrt{m}\,|\hat{\tau}_{\mathtt{DIM}}(\mathtt{t}_S) - \tau| = \sqrt{\mathscr{I}^{\mathtt{DIM}}(\hat{\imath}_{\mathtt{DIM}}, \hat{\jmath}_{\mathtt{DIM}})} + o_{\mathbb{P}}(1) = \sigma\sqrt{(2 + o_{\mathbb{P}}(1))\log\log n}.$$

Theorem SA-18, summed over coordinates and both boundaries and then with $a \downarrow 0$, gives

$$\liminf_{n\to\infty} \mathbb{P}(\mathcal{V}_{\mathtt{DIM}} \neq \varnothing,\ m \le n^b) \ge b.$$

To obtain a deterministic threshold, put $L_n = \log\log n$ and define

$$D_n = \mathbb{1}(\mathcal{V}_{\mathtt{DIM}} \neq \varnothing,\ m \le n^b)\frac{\left|\sqrt{m}\,|\hat{\tau}_{\mathtt{DIM}}(\mathtt{t}_S) - \tau| - \sqrt{\mathscr{I}^{\mathtt{DIM}}(\hat{\imath}_{\mathtt{DIM}}, \hat{\jmath}_{\mathtt{DIM}})}\right|}{\sigma\sqrt{L_n}},$$

with $D_n = 0$ outside the event in the indicator. The preceding bounds give $D_n = o_{\mathbb{P}}(1)$. Thus there is a deterministic sequence $\delta_n \downarrow 0$ such that, with probability tending to one, simultaneously $D_n \le \delta_n$ and

$$\max_{(k,\ell)\in\mathcal{V}_{\mathtt{DIM}}} \mathscr{I}^{\mathtt{DIM}}(k,\ell) \ge \sigma^2(2 - \delta_n)L_n.$$

On this event and $m \le n^b$,

$$|\hat{\tau}_{\mathtt{DIM}}(\mathtt{t}_S) - \tau| \ge \sigma n^{-b/2}\sqrt{L_n}\left(\sqrt{2 - \delta_n} - \delta_n\right).$$

The final factor has square $2 + o(1)$ and is positive for all sufficiently large n, giving the deterministic $o(1)$ in the boundary lower bound.

For the pointwise assertion, put $j = \lceil n^a \rceil$ and $t_n = \eta_n n^{a-1}$. The two threshold variables have laws

$$x_{\pi_\ell(j),\ell} \sim \mathsf{Beta}(j, n+1-j), \qquad 1 - x_{\pi_\ell(n-j),\ell} \sim \mathsf{Beta}(j+1, n-j).$$

For either variable U_n, (SA-31) gives

$$\frac{\mathbb{V}[U_n]}{(\mathbb{E}[U_n])^2} = O(n^{-a}), \qquad \frac{t_n}{\mathbb{E}[U_n]} \to 0, \qquad \mathbb{P}(U_n \le t_n) \le \frac{\mathbb{V}[U_n]}{(\mathbb{E}[U_n] - t_n)^2} = o(1).$$

The $j+1$ in the right-boundary law follows from the largest-left-observation threshold convention. Thus, except on an event of probability $o(1)$,

$$\begin{aligned}\{n^a \le \hat{\imath}_{\mathtt{DIM}} \le n^b,\ \hat{\jmath}_{\mathtt{DIM}} = \ell\} &\implies \{\mathbf{x} : x_\ell \le t_n\} \subseteq \mathtt{t}_S, \\ \{n^a \le n - \hat{\imath}_{\mathtt{DIM}} \le n^b,\ \hat{\jmath}_{\mathtt{DIM}} = \ell\} &\implies \{\mathbf{x} : 1 - x_\ell \le t_n\} \subseteq \mathtt{t}_S.\end{aligned}$$

Each rank event has probability $(b-a)/(2p) + o(1)$ by the split-location limit. Intersect it with the threshold and error events by a union bound. Every point of $\mathcal{X}_n(a, \eta_n)$ lies in at least one strip, so the same finite minimum argument used in Theorem SA-2 proves the uniform infimum bound.

SA-5.21. Proof of Theorem SA-20. Recall the centered DIM contrast $\Delta(\mathsf{t})$ defined in Section SA-3. Under Assumption SA-2, $\tau = \mu_1 - \mu_0$. Hence, for every valid terminal node t,

$$\hat{\tau}_{\mathtt{DIM}}(\mathsf{t}; \mathsf{T}, \mathcal{D}) - \tau = \Delta(\mathsf{t}).$$

Use the centered scalar proxy for the root criterion,

$$\bar{\tau}^{\mathtt{IPW}}(\mathsf{t}) = \frac{1}{n(\mathsf{t})} \sum_{i:\mathbf{x}_i \in \mathsf{t}} \left\{ \frac{d_i \varepsilon_i(1)}{\xi} - \frac{(1-d_i)\varepsilon_i(0)}{1-\xi} \right\}.$$

These proxy averages aggregate exactly under refinement, whereas the arm contrast need not.

Lemma SA-48 (Uniform rectangle control)**.** *Suppose Assumption SA-2 holds and p is fixed. Let $\mathcal{R}_n$ be the finite collection of all axis-aligned sample rectangles whose coordinate endpoints are chosen from the observed coordinate values and the support boundaries. There is an event $\mathcal{G}_n$ with $\mathbb{P}(\mathcal{G}_n^c) = O(n^{-6})$ on which, simultaneously for all $\mathsf{t} \in \mathcal{R}_n$,*

$$|n_1(\mathsf{t}) - \xi n(\mathsf{t})| \le C\sqrt{n(\mathsf{t}) \log n},$$
$$\left| \sum_{i:\mathbf{x}_i \in \mathsf{t}} \mathbb{1}(d_i = d)\varepsilon_i(d) \right| \le C\left\{ \sqrt{n_d(\mathsf{t}) \log n} + \log n \right\}, \qquad d = 0, 1.$$

On the same event, every rectangle with both arms present satisfies

$$|\bar{\tau}^{\mathtt{IPW}}(\mathsf{t})| \le |\Delta(\mathsf{t})| + \frac{C \log n}{n(\mathsf{t})}. \tag{SA-68}$$

For $n(\mathsf{t}) \ge C_0 \log n$, where C_0 is a sufficiently large fixed constant,

$$n(\mathsf{t})|\Delta(\mathsf{t}) - \bar{\tau}^{\mathtt{IPW}}(\mathsf{t})| \le C \log n. \tag{SA-69}$$

Proof. Given the covariates, a coordinate interval selects a consecutive block in its coordinate ordering. Two of the $n+1$ cut positions determine the block, so

$$\#\left\{ \{i : \mathbf{x}_i \in \mathsf{t}\} : \mathsf{t} \in \mathcal{R}_n \right\} \le (n+1)^{2p} \le (n+2)^{2p}.$$

For a fixed membership set containing r observations, the treatment count is a sum of r independent Bernoulli(ξ) assignments. The bounded-variable version of (SA-18) then gives

$$\mathbb{P}\left(|n_1(\mathsf{t}) - \xi r| > C\sqrt{r \log n} \,\middle|\, \mathbf{X} \right) \le 2n^{-cC^2}$$

when the displayed threshold is at most r. Indeed, with $x = C\sqrt{r \log n} \le r$, the Bernstein exponent satisfies $\min\{x^2/r, x\} = C^2 \log n$. If $x > r$, the event is empty. A union bound over all rectangles then makes the count assertion simultaneous.

Conditional on covariates and assignments, each arm sum consists of $n_d(\mathsf{t})$ independent centered observations from the fixed arm law. The exponential-moment assumption in turn permits (SA-18), giving, for $t > 0$,

$$\mathbb{P}\left(\left| \sum_{i:\mathbf{x}_i \in \mathsf{t}} \mathbb{1}(d_i = d)\varepsilon_i(d) \right| > C\{\sqrt{n_d(\mathsf{t})t} + t\} \,\middle|\, \mathbf{X}, \mathbf{d} \right) \le 2e^{-t}.$$

Take $t = (2p+9)\log n$ and sum over the two arms and all rectangle membership sets. Since their number is at most $(n+2)^{2p}$, increasing a fixed constant C gives

$$\mathcal{G}_n = \bigcap_{\mathsf{t} \in \mathcal{R}_n} \left\{ |n_1(\mathsf{t}) - \xi n(\mathsf{t})| \le C\sqrt{n(\mathsf{t}) \log n} \right\} \cap \bigcap_{\substack{\mathsf{t} \in \mathcal{R}_n \\ d \in \{0,1\}}} \left\{ \left| \sum_{i:\mathbf{x}_i \in \mathsf{t}} \mathbb{1}(d_i = d)\varepsilon_i(d) \right| \le C\left(\sqrt{n_d(\mathsf{t}) \log n} + \log n \right) \right\},$$

$\mathbb{P}(\mathcal{G}_n^c) = O(n^{-6})$.

This event is defined before any rectangle is selected using outcomes.

Put $r = n(\mathsf{t})$ and $\hat{\xi}_{\mathsf{t}} = n_1(\mathsf{t})/r$. For every valid rectangle,

$$\Delta(\mathsf{t}) - \bar{\tau}^{\mathtt{IPW}}(\mathsf{t}) = (\xi - \hat{\xi}_{\mathsf{t}}) \left\{ \frac{\bar{\varepsilon}_{1,\mathsf{t}}}{\xi} + \frac{\bar{\varepsilon}_{0,\mathsf{t}}}{1-\xi} \right\}.$$

If $r \geq C_0 \log n$, choose C_0 sufficiently large that $C\sqrt{\log n/r} \leq \min\{\xi, 1-\xi\}/2$. The count bound then gives $n_d(\mathtt{t}) \geq \xi_d r/2$. Hence

$$|\bar{\varepsilon}_{d,\mathtt{t}}| \leq C\left\{\sqrt{\frac{\log n}{r}} + \frac{\log n}{r}\right\} \leq C\sqrt{\frac{\log n}{r}}, \qquad |\hat{\xi}_{\mathtt{t}} - \xi| \leq C\sqrt{\frac{\log n}{r}}.$$

Substitution in the identity then proves $|\Delta(\mathtt{t}) - \bar{\tau}^{\mathtt{IPW}}(\mathtt{t})| \leq C \log n/r$, which is (SA-69). For (SA-68), choose

$$d_\star = \mathbb{1}(\hat{\xi}_{\mathtt{t}} > \xi), \qquad w_{\mathtt{t}} = \min\left\{\frac{\hat{\xi}_{\mathtt{t}}}{\xi}, \frac{1-\hat{\xi}_{\mathtt{t}}}{1-\xi}\right\} \in [0,1].$$

This arm satisfies $n_{d_\star}(\mathtt{t}) \geq \min\{\xi, 1-\xi\}r$, and direct expansion gives

$$\bar{\tau}^{\mathtt{IPW}}(\mathtt{t}) = w_{\mathtt{t}}\Delta(\mathtt{t}) + \frac{\hat{\xi}_{\mathtt{t}} - \xi}{\xi(1-\xi)}\bar{\varepsilon}_{d_\star,\mathtt{t}}. \tag{SA-70}$$

On $\mathcal{G}_n$,

$$r|\hat{\xi}_{\mathtt{t}} - \xi| \leq \min\{r, C\sqrt{r\log n}\}, \qquad |\bar{\varepsilon}_{d_\star,\mathtt{t}}| \leq C\left(\sqrt{\frac{\log n}{r}} + \frac{\log n}{r}\right).$$

Consequently,

$$\begin{aligned} r|\hat{\xi}_{\mathtt{t}} - \xi||\bar{\varepsilon}_{d_\star,\mathtt{t}}| &\leq C\begin{cases}\sqrt{r\log n} + \log n, & r \leq \log n,\\ \log n + (\log n)^{3/2}/\sqrt{r}, & r > \log n\end{cases}\\ &\leq C\log n. \end{aligned}$$

Use $0 \leq w_{\mathtt{t}} \leq 1$ in (SA-70) to obtain (SA-68). □

Lemma SA-49 (Refinement transfer)**.** *On $\mathcal{G}_n$, for every valid parent rectangle $\mathtt{t}$ of size $M \geq C_0 \log n$ and every partition $\mathcal{P}(\mathtt{t})$ into J valid sample rectangles,*

$$\max_{\mathtt{t}' \in \mathcal{P}(\mathtt{t})} |\Delta(\mathtt{t}')| \geq |\Delta(\mathtt{t})| - \frac{CJ\log n}{M}. \tag{SA-71}$$

Proof. The proxy is an ordinary average, so

$$\bar{\tau}^{\mathtt{IPW}}(\mathtt{t}) = \sum_{\mathtt{t}' \in \mathcal{P}(\mathtt{t})} \frac{n(\mathtt{t}')}{M}\bar{\tau}^{\mathtt{IPW}}(\mathtt{t}').$$

For $M \geq C_0 \log n$, (SA-69) and (SA-68) give

$$\begin{aligned} |\Delta(\mathtt{t})| &\leq |\bar{\tau}^{\mathtt{IPW}}(\mathtt{t})| + \frac{C\log n}{M}\\ &\leq \sum_{\mathtt{t}' \in \mathcal{P}(\mathtt{t})} \frac{n(\mathtt{t}')}{M}\left\{|\Delta(\mathtt{t}')| + \frac{C\log n}{n(\mathtt{t}')}\right\} + \frac{C\log n}{M}\\ &\leq \max_{\mathtt{t}' \in \mathcal{P}(\mathtt{t})} |\Delta(\mathtt{t}')| + \frac{C(J+1)\log n}{M}. \end{aligned}$$

Since $J + 1 \leq 2J$, this proves (SA-71). All rectangle bounds hold simultaneously, so the parent and partition may depend on the outcomes. □

Corollary SA-50 (Terminal inheritance from a root child with random size)**.** *Fix $0 < a < b < 1$, and let J_n be deterministic with*

$$J_n \log n = o\left(n^{a/2}\sqrt{\log\log n}\right).$$

Suppose that on an event $\mathcal{E}_n$ a valid sample rectangle $\mathtt{t}_n$ and its partition $\mathcal{P}(\mathtt{t}_n)$ into at most J_n valid terminal descendants satisfy

$$M_n = n(\mathtt{t}_n) \in [n^a, n^b], \qquad |\Delta(\mathtt{t}_n)| \geq c_{\mathrm{root}} M_n^{-1/2}\sqrt{\log\log n}$$

for some constant $c_{\mathrm{root}} > 0$. These node conditions are required only on $\mathcal{E}_n$. Then

$$\mathbb{P}\left(\mathcal{E}_n, \max_{\mathtt{t}' \in \mathcal{P}(\mathtt{t}_n)} |\Delta(\mathtt{t}')| \geq \frac{c_{\mathrm{root}}}{2} M_n^{-1/2}\sqrt{\log\log n}\right) \geq \mathbb{P}(\mathcal{E}_n) - o(1).$$

Consequently,

$$\mathbb{P}\left(\mathcal{E}_n, \max_{\mathtt{t}' \in \mathcal{P}(\mathtt{t}_n)} |\Delta(\mathtt{t}')| \geq \frac{c_{\mathrm{root}}}{2} n^{-b/2}\sqrt{\log\log n}\right) \geq \mathbb{P}(\mathcal{E}_n) - o(1).$$

Proof. On $\mathcal{E}_n \cap \mathcal{G}_n$, we have $M_n \geq n^a \geq C_0 \log n$ for all sufficiently large n. Equation (SA-71) then gives

$$\begin{aligned}\max_{\mathtt{t}' \in \mathcal{P}(\mathtt{t}_n)} |\Delta(\mathtt{t}')| &\geq c_{\mathrm{root}} M_n^{-1/2} \sqrt{\log\log n} - \frac{C J_n \log n}{M_n} \\ &\geq \left(c_{\mathrm{root}} - \frac{C J_n \log n}{n^{a/2}\sqrt{\log\log n}} \right) M_n^{-1/2} \sqrt{\log\log n} \\ &\geq \frac{c_{\mathrm{root}}}{2} M_n^{-1/2} \sqrt{\log\log n}.\end{aligned}$$

Thus the first event in the conclusion contains $\mathcal{E}_n \cap \mathcal{G}_n$ for all large n, and

$$\mathbb{P}(\mathcal{E}_n \cap \mathcal{G}_n) \geq \mathbb{P}(\mathcal{E}_n) - \mathbb{P}(\mathcal{G}_n^c) = \mathbb{P}(\mathcal{E}_n) - o(1).$$

The second assertion in turn follows from $M_n^{-1/2} \geq n^{-b/2}$. □

To prove Theorem SA-20, fix $0 < a < b < 1$ with $2^K \log n = o(n^{a/2}\sqrt{\log\log n})$. On $\mathcal{V}_{\mathtt{DIM}} \neq \varnothing$, let $\mathtt{t}_n$ be the smaller selected root child (the left child in a tie), and put $M_n = n(\mathtt{t}_n)$. Define

$$\mathcal{E}_{\mathtt{DIM},n} = \left\{ \mathcal{V}_{\mathtt{DIM}} \neq \varnothing, \quad n^a \leq M_n \leq n^b, \quad \sqrt{M_n}|\Delta(\mathtt{t}_n)| \geq c_{\mathrm{root}} \sqrt{\log\log n} \right\}.$$

The selected-child conditions apply only when the root is split. The normalized smaller-child error bound in the proof of Theorem SA-19 and the split-location limits in Theorem SA-18 together give the required error and size conditions. Combining the left and reflected right boundary events gives constants $c_{\mathrm{root}}, q_{\mathrm{root}} > 0$, depending only on the distributional constants in Assumption SA-2 and the fixed dimension p, such that

$$\liminf_{n\to\infty} \mathbb{P}(\mathcal{E}_{\mathtt{DIM},n}) \geq q_{\mathrm{root}}(b-a).$$

For DIM, take $c_{\mathrm{root}} = \sigma$ and $q_{\mathrm{root}} = 1$, since the normalized selected-child error tends to $\sqrt{2}\sigma$ and the two rank events have combined limiting probability $b - a$.

Let $\mathcal{P}(\mathtt{t}_n)$ be the terminal descendants of $\mathtt{t}_n$. Valid recursive splitting preserves both arms in every descendant. On $\mathcal{E}_{\mathtt{DIM},n}$,

$$M_n \in [n^a, n^b], \qquad |\mathcal{P}(\mathtt{t}_n)| \leq 2^K, \qquad \frac{2^K \log n}{n^{a/2}\sqrt{\log\log n}} \to 0.$$

Applying Corollary SA-50 then gives

$$\mathbb{P}\left(\mathcal{E}_{\mathtt{DIM},n}, \max_{\mathtt{t}' \in \mathcal{P}(\mathtt{t}_n)} |\Delta(\mathtt{t}')| \geq \frac{c_{\mathrm{root}}}{2} n^{-b/2} \sqrt{\log\log n} \right) \geq \mathbb{P}(\mathcal{E}_{\mathtt{DIM},n}) - o(1).$$

For every terminal descendant $\mathtt{t}'$ and every $\mathbf{x} \in \mathtt{t}'$,

$$\hat{\tau}_{\mathtt{DIM}}(\mathbf{x}) - \tau(\mathbf{x}) = \Delta(\mathtt{t}').$$

Thus the supremum over $\mathbf{x} \in \mathcal{X}$ is at least the maximum over these terminal descendants. Consequently,

$$\liminf_{n\to\infty} \mathbb{P}\left(\sup_{\mathbf{x}\in\mathcal{X}} |\hat{\tau}_{\mathtt{DIM}}(\mathbf{x}) - \tau(\mathbf{x})| \geq \frac{c_{\mathrm{root}}}{2} n^{-b/2} \sqrt{\log\log n} \right) \geq q_{\mathrm{root}}(b-a).$$

Using $c_{\mathrm{root}} = \sigma$ and $q_{\mathrm{root}} = 1$ proves (SA-11). Taking $a = b/2$ then gives its stated specialization.

For (SA-12), let $2\rho_K < b$ and fix $a \in (2\rho_K, b)$. Put $\delta = (a/2 - \rho_K)/2 > 0$. By the definition of ρ_K, eventually

$$2^{K_n} \leq n^{\rho_K + \delta}, \qquad \frac{2^{K_n} \log n}{n^{a/2}\sqrt{\log\log n}} \leq n^{-\delta} \frac{\log n}{\sqrt{\log\log n}} \to 0. \qquad \text{[SA-72]}$$

Equation (SA-11) therefore applies for every fixed $a \in (2\rho_K, b)$. Its error event and constants do not depend on a, so

$$\liminf_{n\to\infty} \mathbb{P}\left(\sup_{\mathbf{x}\in\mathcal{X}} |\hat{\tau}_{\mathtt{DIM}}(\mathbf{x}) - \tau(\mathbf{x})| \geq \frac{\sigma}{2} n^{-b/2} \sqrt{\log\log n} \right) \geq \sup_{a \in (2\rho_K, b)} (b - a) = b - 2\rho_K.$$

This takes the supremum over fixed a after the limit inferior in n.

SA-5.22. Proof of Theorem SA-21. Abbreviate $\hat{\tau}_{\mathtt{DIM}}$ by $\hat{\tau}$. We use the rectangle argument from the proof of Theorem SA-4, now separately for each treatment arm. For a deterministic grid rectangle Q and $d \in \{0,1\}$,

$$n_d(Q) \sim \mathsf{Binomial}\big(n, \xi_d P_X(Q)\big).$$

Use open rectangles with endpoints in the deterministic coordinate grid of mesh n^{-2} on the normalized cube, together with the empty rectangle. There are at most $(n^2+1)^{2p}+1$ such rectangles. For a grid rectangle Q, the binomial lower-tail estimate (SA-19) gives

$$\mathbb{P}\left(n_d(Q) \le \tfrac{1}{2} n\xi_d P_X(Q)\right) \le \exp\{-n\xi_d P_X(Q)/8\}.$$

If $n\xi_d P_X(Q) \ge A \log n$, the right side is at most $n^{-A/8}$. Choose $A > 8(4p+8)$ and take a union bound over both arms and the grid rectangles. Rectangles with $n\xi_d P_X(Q) < A\log n$ already satisfy the desired bound deterministically. Thus, except on an event of probability $O(n^{-6})$,

$$P_X(Q) \le \frac{2n_d(Q) + A\log n}{n\xi_d}$$

for every grid rectangle and both arms. For an arbitrary rectangle t, round each lower endpoint up and each upper endpoint down, and take the open rectangle between these rounded endpoints. The inclusion $Q \subseteq \mathsf{t}$ holds, by construction, for every endpoint convention for t. By independence and uniformity of the normalized coordinates, $P_X(\mathsf{t} \setminus Q) \le 2pn^{-2}$. The same bound holds if the inner rectangle is empty because some side is shorter than two grid spacings. Since $n_d(Q) \le n_d(\mathsf{t})$ and ξ_d is fixed and positive, we obtain simultaneously

$$P_X(\mathsf{t}) \le C\frac{n_d(\mathsf{t}) + \log n}{n\xi_d}.$$

Conditional on all covariates and treatment assignments, the errors within each arm remain independent draws from their original centered subexponential law. Intersecting rectangle membership subsets with the fixed assignment vector does not increase their number. Apply (SA-18) with $t = (2p+7)\log n$ to each such subset and each arm. Its sum bound, divided by the positive count $n_d(\mathsf{t})$, is

$$C\left\{\sqrt{\frac{\log n}{n_d(\mathsf{t})}} + \frac{\log n}{n_d(\mathsf{t})}\right\}.$$

Its square is at most $C\log^2 n/n_d(\mathsf{t})$, because $n_d(\mathsf{t}) \ge 1$ and $n \ge 2$, with a change in the fixed constant. Write $\mathcal{R}$ for all axis-aligned rectangles in the normalized cube, and fix C sufficiently large. The union bound over at most $2(n+2)^{2p}$ distinct sample membership sets, together with the grid argument, therefore gives

$$\mathcal{E}_n = \bigcap_{\substack{\mathsf{t}\in\mathcal{R}\\ d\in\{0,1\}}} \left\{P_X(\mathsf{t}) \le C\frac{n_d(\mathsf{t})+\log n}{n\xi_d}\right\} \cap \bigcap_{\substack{\mathsf{t}\in\mathcal{R},\ d\in\{0,1\}\\ n_d(\mathsf{t})>0}} \left\{\left|\frac{1}{n_d(\mathsf{t})}\sum_{i:\mathbf{x}_i\in\mathsf{t},\ d_i=d}\varepsilon_i(d)\right|^2 \le C\frac{\log^2 n}{n_d(\mathsf{t})}\right\}, \qquad \mathbb{P}(\mathcal{E}_n^c) \le Cn^{-6}.$$

For measurability, the residual bounds need only be checked on finitely many sample membership sets. The mass bounds can, in turn, be checked on rectangles with rational endpoints and then extended by one-sided limits. The constants depend only on q, the distribution of $(y_i(0), y_i(1), d_i)$, and the fixed dimension p.

First suppose $q \ge 2$. On a leaf with both arm counts positive, $\hat{\tau}(\mathsf{t}) - \tau$ is the difference of the two arm residual averages. Using $|a-b|^q \le 2^{q-1}(|a|^q + |b|^q)$ and the mass bound for each arm gives, on $\mathcal{E}_n$,

$$P_X(\mathsf{t})|\hat{\tau}(\mathsf{t}) - \tau|^q \le \frac{C\log^q n}{n}\sum_{d=0}^{1}\frac{n_d(\mathsf{t})^{1-q/2} + \log n\, n_d(\mathsf{t})^{-q/2}}{\xi_d} \le C\frac{\log^{q+1} n}{n}.$$

If $n_d(\mathsf{t}) = 0$ for either arm, the entire DIM terminal estimate is zero, so its contribution satisfies

$$P_X(\mathsf{t})|\tau|^q \le C\frac{|\tau|^q \log n}{n\xi_d} \le C\frac{\log^{q+1} n}{n}.$$

This also covers leaves with no sample observations. Summing over at most 2^K leaves then bounds the integrated q-th-power error on $\mathcal{E}_n$ by $C2^K \log^{q+1} n/n$.

For every sample, the integrated q-th-power error is at most

$$|\tau|^q + 2^q \max_{1\le i\le n,\ d\in\{0,1\}} |\varepsilon_i(d)|^q.$$

Write $H_n = |\tau|^q + 2^q \max_{1\le i\le n,\ d\in\{0,1\}} |\varepsilon_i(d)|^q$. The finite $2q$-th moments in turn give

$$\mathbb{E}[H_n^2] \le C\left\{ |\tau|^{2q} + \sum_{i=1}^{n}\sum_{d=0}^{1} \mathbb{E}[|\varepsilon_i(d)|^{2q}] \right\} \le Cn.$$

Thus

$$\mathbb{E}\big[\|\hat\tau - \tau\|_q^q \mathbb{1}(\mathcal{E}_n^c)\big] \le \{\mathbb{E}[H_n^2]\mathbb{P}(\mathcal{E}_n^c)\}^{1/2} \le Cn^{-5/2}.$$

Together with the bound on $\mathcal{E}_n$, this gives

$$\mathbb{E}[\|\hat\tau - \tau\|_q^q] \le C\frac{2^K \log^{q+1} n}{n}, \qquad q \ge 2.$$

For $1 \le q < 2$, Hölder's and Jensen's inequalities give

$$\|\hat\tau - \tau\|_q \le \|\hat\tau - \tau\|_2, \qquad \mathbb{E}[\|\hat\tau - \tau\|_q^q] \le \big(\mathbb{E}[\|\hat\tau - \tau\|_2^2]\big)^{q/2} \le C\left(\frac{2^K \log^3 n}{n}\right)^{q/2}.$$

Taking q-th roots then proves the moment bounds. The corresponding inequalities on $\mathcal{E}_n$ yield, for a sufficiently large fixed C_2,

$$\left\{ \|\hat\tau - \tau\|_q \ge C_2\left(\frac{2^K \log^{\max\{3,q+1\}} n}{n}\right)^{1/\max\{2,q\}} \right\} \subseteq \mathcal{E}_n^c, \qquad \mathbb{P}(\mathcal{E}_n^c) \le Cn^{-6}.$$

The rectangle bounds hold simultaneously and 2^K is deterministic, so the partition may depend on the outcomes.

SA-5.23. Proof of Theorem SA-22. We first give the conditional anti-concentration bound used for the selected terminal node.

Lemma SA-51 (Independent finite-arm anti-concentration)**.** *Suppose Assumption SA-2 holds. Let $\mathtt{t}_\star$ be a terminal node selected using only a construction fold tree and estimation fold covariates, not estimation fold treatment assignments or outcomes. Conditional on $\mathtt{t}_\star$ and on treatment counts $N_{1,\star}, N_{0,\star}$ satisfying $N_{1,\star} > 0$ and $N_{0,\star} > 0$, the honest DIM terminal contrast obeys*

$$\mathbb{P}\left(|\check\tau_{\mathtt{DIM}}(\mathtt{t}_\star) - \tau| \ge C_1 N_\star^{-1/2} \,\middle|\, \mathtt{t}_\star,\ N_{1,\star},\ N_{0,\star} \right) \ge C_2, \qquad N_\star = N_{1,\star} + N_{0,\star},$$

for constants $C_1, C_2 > 0$ depending only on the distribution of $(y_i(0), y_i(1), d_i)$. The same bound holds conditional on the construction data and all covariates and treatment assignments in the estimation fold.

Proof. Condition first on the construction data, all estimation covariates, and all estimation assignments. Write $\mathcal{F}$ for this sigma-field and $\varphi_d(t) = \mathbb{E}[e^{it\varepsilon_i(d)}]$. The selected indices are $\mathcal{F}$-measurable. Independence of their residuals then gives, when $N_{1,\star}N_{0,\star} > 0$,

$$\mathbb{E}\left[e^{it\{\check\tau_{\mathtt{DIM}}(\mathtt{t}_\star) - \tau\}} \,\middle|\, \mathcal{F} \right] = \varphi_1(t/N_{1,\star})^{N_{1,\star}} \varphi_0(-t/N_{0,\star})^{N_{0,\star}}.$$

The right side depends only on the two counts. Iterated conditioning and uniqueness of characteristic functions [18, Theorem 3.3.11] therefore give the same contrast law conditional on $\mathtt{t}_\star, N_{1,\star}, N_{0,\star}$. Its variance is

$$v = \frac{\sigma_1^2}{N_{1,\star}} + \frac{\sigma_0^2}{N_{0,\star}} \ge \frac{c_\varepsilon}{N_\star}, \qquad \sigma_d^2 = \mathbb{V}[\varepsilon_i(d)] > 0, \quad c_\varepsilon = \sigma_0^2 + \sigma_1^2.$$

To verify the uniformity in the counts, let X be this centered contrast and let $\kappa_4 = \max_{d=0,1} \mathbb{E}[\varepsilon_i(d)^4]/\sigma_d^4 < \infty$. Expanding the fourth power then gives

$$\mathbb{E}\big[X^4 \mid \mathtt{t}_\star,\ N_{1,\star},\ N_{0,\star} \big] = 3v^2 + \sum_{d=0}^{1} \frac{\mathbb{E}[\varepsilon_i(d)^4] - 3\sigma_d^4}{N_{d,\star}^3} \le 3v^2 + (\kappa_4 - 3)_+ \sum_{d=0}^{1} \frac{\sigma_d^4}{N_{d,\star}^3} \le \max\{3, \kappa_4\}\, v^2,$$

since $N_{d,\star} \ge 1$ and $\sum_{d=0}^1 \sigma_d^4/N_{d,\star}^3 \le \sum_{d=0}^1 (\sigma_d^2/N_{d,\star})^2 \le v^2$. Apply the Paley–Zygmund calculation in the proof of Lemma SA-42 to X^2 at half its conditional mean. It then yields

$$\mathbb{P}\left(|X| \ge \sqrt{v/2} \,\middle|\, \mathtt{t}_\star,\ N_{1,\star},\ N_{0,\star} \right) \ge \frac{1}{4\max\{3, \kappa_4\}}.$$

Since $v \ge c_\varepsilon/N_\star$, take $C_1 = \sqrt{c_\varepsilon/2}$ and $C_2 = 1/\{4\max\{3,\kappa_4\}\}$. We have thus applied the argument of Lemma SA-42 with the two sets of arm weights. The arm samples contain distinct individuals, so their independence does not require the two potential errors of any individual to be independent. □

To apply this bound, we next control the estimation counts in a selected terminal node. Let $M = n_{\mathsf{T}}$ and $N = n_\tau$, and let $\tilde{\mathbf{X}}$ denote the matrix of estimation-fold covariates. Fix $b \in (0,1)$ and set $a = b/2$. Theorem SA-18, summed over coordinates, gives an event

$$\mathcal{E}_n = \{\mathcal{V}_{\texttt{DIM}}(M) \neq \varnothing, \quad M^a \leq \hat{\imath}_{\texttt{DIM}} \leq M^b\}$$

with $\liminf \mathbb{P}(\mathcal{E}_n) \geq b/4$. On this event, let t_{L} be the selected left root child, with construction size $m = \hat{\imath}_{\texttt{DIM}}$, and let R be its estimation count. The selected root, as in the regression argument, depends only on the coordinate rank permutations, construction assignments, and potential errors. By (SA-30), these are independent of the coordinate order-statistic vectors. Thus, given the selected coordinate and rank m, the population mass W of the left child has law $\mathsf{Beta}(m, M+1-m)$. Given W, the independent estimation count has law $\mathsf{Binomial}(N, W)$. Using the beta moments and total variance, as in (SA-32), then gives

$$\mathbb{E}\big[\, R \mid \hat{\jmath}_{\texttt{DIM}},\, \hat{\imath}_{\texttt{DIM}} = m \,\big] = \frac{Nm}{M+1}, \qquad \mathbb{V}\big[\, R \mid \hat{\jmath}_{\texttt{DIM}},\, \hat{\imath}_{\texttt{DIM}} = m \,\big] = \frac{Nm(M+1-m)(M+1+N)}{(M+1)^2(M+2)} \leq Cm.$$

Take $c_R = \rho/4$ and $C_R = 2/\rho$. The hypothesis $\rho \leq M/N \leq \rho^{-1}$ then gives

$$2c_R m \leq \mathbb{E}[R \mid \hat{\jmath}_{\texttt{DIM}},\, \hat{\imath}_{\texttt{DIM}} = m] \leq \tfrac{1}{2} C_R m.$$

Chebyshev's inequality therefore gives, uniformly over the selected coordinate and $m \in [M^a, M^b]$,

$$\mathbb{P}\left(R \notin [c_R m, C_R m] \mid \hat{\jmath}_{\texttt{DIM}},\, \hat{\imath}_{\texttt{DIM}} = m\right) \leq \frac{C}{m} \leq C M^{-a}.$$

Averaging over $\mathcal{E}_n$ in turn gives

$$\mathbb{P}(c_R m \leq R \leq C_R m \mid \mathcal{E}_n) \to 1, \qquad R \leq C_R n^b \quad \text{on this count event.}$$

Order the terminal descendants $\mathsf{t}_1, \ldots, \mathsf{t}_J$ of t_{L} by their left-to-right tree addresses, and let N_j be their estimation-fold counts. Valid recursive splitting ensures that each contains both construction arms, so $J \leq m/2$. Let $Z_1, \ldots, Z_R$ be the descendant labels of the estimation observations in this root child, in their original sample order, and put $u_j = P_X(\mathsf{t}_j)/P_X(\mathsf{t}_{\mathsf{L}})$. Conditional on the construction fold and the indices falling in the root child, independence gives a product of these label probabilities. This product does not depend on the indices, so

$$\mathbb{P}\left(\, Z_1 = z_1,\ \ldots,\ Z_R = z_R \,\middle|\, \mathcal{D}_{\mathsf{T}},\, R \,\right) = \prod_{h=1}^{R} u_{z_h}, \qquad z_h \in [J].$$

Pair these labels into $\lfloor R/2 \rfloor$ disjoint pairs. The product law and $1 = (\sum_{j=1}^J u_j)^2 \leq J \sum_{j=1}^J u_j^2$ together give

$$\mathbb{P}\left(\max_{1 \leq j \leq J} N_j \geq 2 \,\middle|\, \mathcal{D}_{\mathsf{T}},\, R \right) \geq 1 - \left(1 - \sum_{j=1}^{J} u_j^2\right)^{\lfloor R/2 \rfloor} \geq 1 - \exp\{-\lfloor R/2 \rfloor / J\}.$$

For all sufficiently large n, on $R \geq c_R m$ and $m \geq M^a$ we have $\lfloor R/2 \rfloor \geq c_R m/3$. Since $J \leq m/2$, the right side is at least $q_{\text{occ}} := 1 - \exp(-2c_R/3) = 1 - e^{-\rho/6} > 0$. Define

$$\mathcal{H}_n = \mathcal{E}_n \cap \{c_R m \leq R \leq C_R m\} \cap \left\{ \max_{1 \leq j \leq J} N_j \geq 2 \right\}.$$

As with m and the descendants, the last two conditions are imposed on $\mathcal{E}_n$. Conditioning first on the construction fold and R proves

$$\mathbb{P}(\mathcal{H}_n) \geq q_{\text{occ}} \mathbb{P}(\mathcal{E}_n,\ c_R m \leq R \leq C_R m), \qquad \liminf_{n \to \infty} \mathbb{P}(\mathcal{H}_n) \geq q_{\text{occ}} b/4.$$

The argument only requires the positive occupancy bound above, not convergence of the occupancy probability to one.

On $\mathcal{H}_n$, put $j_\star = \min\{j : N_j \geq 2\}$, $\mathsf{t}_\star = \mathsf{t}_{j_\star}$, and $N_\star = N_{j_\star}$. Choose the leftmost terminal node outside $\mathcal{H}_n$. Both this selection and $\mathcal{H}_n$ are measurable with respect to $\mathcal{F}_n = \sigma(\mathcal{D}_{\mathsf{T}}, \tilde{\mathbf{X}})$, and on $\mathcal{H}_n$,

$$2 \leq N_\star \leq R \leq C_R m \leq C_R n^b.$$

Let $N_{d,\star}$ be the selected node's estimation count in arm d. Conditional on $\mathcal{F}_n$, its $N_\star$ assignments are i.i.d. Bernoulli(ξ) and remain independent of the potential errors. Therefore, on $\mathcal{H}_n$,

$$\mathbb{P}\big(\, N_{1,\star} > 0,\ N_{0,\star} > 0 \mid \mathcal{F}_n \,\big) = 1 - \xi^{N_\star} - (1-\xi)^{N_\star} \geq 2\xi(1-\xi).$$

When both arms are present, apply Lemma SA-51 to the terminal residual contrast, with $c_{\mathrm{ac}} = C_1$ and $q_{\mathrm{ac}} = C_2$ as specified in its proof. We may further condition on $\mathcal{F}_n$ and all estimation assignments because these determine which observations enter the contrast without revealing their errors. Hence

$$\mathbb{P}\left(\sup_{\mathbf{x}\in\mathcal{X}} |\check{\tau}_{\mathtt{DIM}}(\mathbf{x}) - \tau(\mathbf{x})| \geq \frac{c_{\mathrm{ac}}}{\sqrt{C_R}} n^{-b/2} \right) \geq \mathbb{E}\left[\mathbb{1}(\mathcal{H}_n) q_{\mathrm{ac}} \mathbb{P}\big(N_{1,\star} > 0,\ N_{0,\star} > 0 \mid \mathcal{F}_n \big) \right] \geq 2\xi(1-\xi) q_{\mathrm{ac}} \mathbb{P}(\mathcal{H}_n).$$

Consequently,

$$\begin{aligned} \liminf_{n\to\infty} \mathbb{P}\left(\sup_{\mathbf{x}\in\mathcal{X}} |\check{\tau}_{\mathtt{DIM}}(\mathbf{x}) - \tau(\mathbf{x})| \geq \frac{\sqrt{\rho(\sigma_0^2+\sigma_1^2)}}{2} n^{-b/2} \right) \\ \geq 2\xi(1-\xi) q_{\mathrm{ac}} \frac{q_{\mathrm{occ}} b}{4} = \frac{\xi(1-\xi)(1-e^{-\rho/6}) b}{8\max\{3, \mathbb{E}[\varepsilon_i(0)^4]/\sigma_0^4, \mathbb{E}[\varepsilon_i(1)^4]/\sigma_1^4\}}. \end{aligned}$$

The lower bound thus uses only events on which both estimation arms are present.

SA-5.24. Proof of Theorem SA-23. We first consider $q = 2$. Condition on the construction fold $\mathcal{D}_{\mathsf{T}}$, and write $\mathcal{P}$ for its terminal partition and $m = n_\tau$. For a leaf t with $u = P_X(\mathsf{t}) > 0$, each treatment-arm count has conditional marginal distribution

$$n_d(\mathsf{t}) \mid \mathcal{D}_{\mathsf{T}} \sim \mathsf{Binomial}\big(m, u\xi_d\big), \qquad d \in \{0,1\}.$$

Write $\sigma_d^2 = \mathbb{E}[\varepsilon_i(d)^2]$. Given the construction fold and the estimation covariates and assignments, a leaf containing both arms has conditional mean squared error $\sigma_1^2/n_1(\mathsf{t}) + \sigma_0^2/n_0(\mathsf{t})$. The averages involve distinct individuals and are therefore independent, even if an individual's potential errors are dependent. If either arm is absent, the DIM estimate is zero, giving squared error τ^2.

Averaging over the estimation covariates and assignments, dropping the other-arm positivity condition from each variance term, and taking a union bound over the missing-arm events gives, for $\mathbf{x} \in \mathsf{t}$,

$$u\mathbb{E}\big[(\check{\tau}_{\mathtt{DIM}}(\mathbf{x}) - \tau(\mathbf{x}))^2 \mid \mathcal{D}_{\mathsf{T}}\big] \leq u \sum_{d=0}^{1} \left\{ \sigma_d^2 \sum_{k=1}^{m} \frac{\mathbb{P}(n_d(\mathsf{t}) = k \mid \mathcal{D}_{\mathsf{T}})}{k} + \tau^2 \mathbb{P}(n_d(\mathsf{t}) = 0 \mid \mathcal{D}_{\mathsf{T}}) \right\}.$$

The two arm counts need not, however, be independent. For $B \sim \mathsf{Binomial}(m, v)$ and $v > 0$,

$$\mathbb{E}\left[\frac{1}{B+1}\right] = \frac{1}{(m+1)v} \sum_{j=1}^{m+1} \binom{m+1}{j} v^j (1-v)^{m+1-j} = \frac{1-(1-v)^{m+1}}{(m+1)v}.$$

Using $1/k \leq 2/(k+1)$ for $k \geq 1$ and $\sup_{0\leq v\leq 1} v(1-v)^m = m^m/(m+1)^{m+1} \leq 1/(m+1)$ with $v = u\xi_d$ yields

$$u \sum_{k=1}^{m} \frac{\mathbb{P}(n_d(\mathsf{t}) = k \mid \mathcal{D}_{\mathsf{T}})}{k} \leq \frac{2}{(m+1)\xi_d} \qquad \text{and} \qquad u\mathbb{P}(n_d(\mathsf{t}) = 0 \mid \mathcal{D}_{\mathsf{T}}) \leq \frac{1}{(m+1)\xi_d}.$$

Zero-mass leaves contribute zero. Summing over the terminal partition, using $\xi_1 = \xi$ and $\xi_0 = 1-\xi$, then gives

$$\mathbb{E}\left[\int_{\mathcal{X}} (\check{\tau}_{\mathtt{DIM}}(\mathbf{x}) - \tau(\mathbf{x}))^2 \, \mathrm{d}F_X(\mathbf{x}) \,\middle|\, \mathcal{D}_{\mathsf{T}} \right] \leq \frac{|\mathcal{P}|}{m+1} \left\{ \frac{2\sigma_1^2 + \tau^2}{\xi} + \frac{2\sigma_0^2 + \tau^2}{1-\xi} \right\}.$$

Since $|\mathcal{P}| \leq 2^K$, unconditioning and taking the square root proves the moment bound for $q = 2$ with $C_1 = \{(2\sigma_1^2+\tau^2)/\xi + (2\sigma_0^2+\tau^2)/(1-\xi)\}^{1/2}$. Applying Markov's inequality to integrated squared error and taking square roots in turn gives its probability bound with $C_2 = C_1$.

For fixed $q > 2$, the exponential-moment assumption and (SA-18) give

$$\mathbb{E}\left[\left| \frac{1}{k} \sum_{i=1}^{k} \varepsilon_i(d) \right|^q \right] \leq C_q(k^{-q/2} + k^{-q}) \leq C_q k^{-q/2}, \qquad k \geq 1.$$

Indeed, for $S_{d,k} = \sum_{i=1}^{k} \varepsilon_i(d)$, the sum tail and $e^{-\min\{a,b\}} \leq e^{-a} + e^{-b}$ give

$$\mathbb{E}[|S_{d,k}|^q] = q\int_0^\infty t^{q-1}\mathbb{P}(|S_{d,k}| > t)\,\mathrm{d}t \leq 2q\int_0^\infty t^{q-1}\{e^{-ct^2/k} + e^{-ct}\}\,\mathrm{d}t = qc^{-q/2}\Gamma(q/2)k^{q/2} + 2qc^{-q}\Gamma(q).$$

Divide by k^q. This is the centered-average moment bound used in the scalar honest proof, now for either arm. Conditional on the covariates and assignments, $|a-b|^q \le 2^{q-1}(|a|^q+|b|^q)$ bounds the moment of the contrast whenever both arms are present. Dropping the other-arm positivity restriction and using $k^{-q/2} \le 1/k$ gives

$$u\mathbb{E}\left[|\check{\tau}_{\mathtt{DIM}}(\mathbf{x})-\tau(\mathbf{x})|^q \mid \mathcal{D}_{\mathsf{T}}\right] \le Cu\sum_{d=0}^{1}\sum_{k=1}^{m}\frac{\mathbb{P}(n_d(\mathbf{t})=k \mid \mathcal{D}_{\mathsf{T}})}{k} + u|\tau|^q\sum_{d=0}^{1}\mathbb{P}(n_d(\mathbf{t})=0 \mid \mathcal{D}_{\mathsf{T}}) \le \frac{C}{m+1},$$

where C depends only on q and the outcome/treatment distribution. Summing over the leaves and unconditioning then bounds the expected integrated q-th-power error. Taking its q-th root, and applying Markov's inequality before taking q-th roots of the probability threshold, prove both conclusions. For $1 \le q < 2$, the probability measure inequality $\|g\|_q \le \|g\|_2$ and Jensen's inequality yield

$$\left(\mathbb{E}[\|\check{\tau}_{\mathtt{DIM}}-\tau\|_q^q]\right)^{1/q} \le \left(\mathbb{E}[\|\check{\tau}_{\mathtt{DIM}}-\tau\|_2^2]\right)^{1/2}.$$

At the stated square-root threshold, an L^q exceedance implies an L^2 exceedance. Both constants may therefore be taken as the $q=2$ value of C_1. The argument thus uses neither independence of the two arm counts nor independence of an individual's potential errors.

SA-5.25. Proof of Lemma SA-24. Fix an arm d and one covariate ordering, and write

$$A_k = \sum_{i=1}^{k}\mathbb{1}(d_{\pi_\ell(i)}=d)\varepsilon_{\pi_\ell(i)}(d), \qquad N_k = \sum_{i=1}^{k}\mathbb{1}(d_{\pi_\ell(i)}=d), \qquad B_k = N_k - \xi_d k.$$

The constant arm mean again cancels from a within-arm contrast. The elementary identity

$$\frac{uv}{u+v}\left(\frac{a}{u}-\frac{b}{v}\right)^2 = \frac{a^2}{u}+\frac{b^2}{v}-\frac{(a+b)^2}{u+v}, \qquad u,v>0,$$

follows by multiplying out the square. Apply it first with $(a,b,u,v)=(A_k, A_n-A_k, N_k, N_n-N_k)$ and then with $(a,b,u,v)=(A_k, A_n-A_k, \xi_d k, \xi_d(n-k))$. The arm's empirical contribution and its proxy are therefore exactly

$$\frac{A_k^2}{N_k}+\frac{(A_n-A_k)^2}{N_n-N_k}-\frac{A_n^2}{N_n}, \qquad \frac{A_k^2}{\xi_d k}+\frac{(A_n-A_k)^2}{\xi_d(n-k)}-\frac{A_n^2}{\xi_d n},$$

respectively. On $N_k \ge \xi_d k/2$,

$$\left|\frac{A_k^2}{N_k}-\frac{A_k^2}{\xi_d k}\right| \le C\frac{|A_k|^2|B_k|}{k^2}.$$

Lemma SA-40 bounds the supremum of the right side over $k \ge r_n$ by $O_{\mathbb{P}}(r_n^{-1/2})$. The two-arm denominator event $\mathcal{A}_n(r_n)$ in (SA-67), defined for the current ordering, has probability tending to one and makes every retained split valid. Reversing the ordering similarly gives the same $O_{\mathbb{P}}(r_n^{-1/2})$ bound for the right-child term. On $\mathcal{A}_n(r_n)$,

$$N_n \ge \xi_d n/2, \qquad \left|\frac{A_n^2}{N_n}-\frac{A_n^2}{\xi_d n}\right| \le C\frac{|A_n|^2|B_n|}{n^2} = O_{\mathbb{P}}(n^{-1/2}) = O_{\mathbb{P}}(r_n^{-1/2}),$$

since $A_n, B_n = O_{\mathbb{P}}(\sqrt{n})$ and a nonempty retained range has $r_n \le n/2$. Adding the three differences, then the two arms, and taking a union bound over p coordinates therefore proves the result. On $\mathcal{A}_n(r_n)^c$, whose probability tends to zero, empirical expressions are evaluated only at valid candidates. An empty retained range requires no argument.

SA-5.26. Proof of Lemma SA-25. Retain the notation from the preceding proof. Conditional on treatment assignments, extend the observed errors in arm d to an infinite i.i.d. residual stream with partial sums $T_d(h)$, so $A_k = T_d(N_k)$. For $k \le s_n$, the nonnegative empirical arm contribution is bounded above by

$$\frac{A_k^2}{N_k}+\frac{(A_n-A_k)^2}{N_n-N_k}.$$

On valid candidates, (SA-23) gives

$$\max_{1\le k\le s_n:\, N_k>0}\frac{A_k^2}{N_k} \le \max_{1\le h\le s_n}\frac{T_d(h)^2}{h} = O_{\mathbb{P}}(\log\log s_n).$$

With probability tending to one, $N_n - N_k \ge \xi_d n/2$ throughout this range. On this event,

$$\max_{1\le k\le s_n}\frac{(A_n-A_k)^2}{N_n-N_k} \le \frac{2\left(|A_n|+\max_{1\le h\le s_n}|T_d(h)|\right)^2}{\xi_d n} = O_{\mathbb{P}}(1),$$

since $A_n = O_{\mathbb{P}}(\sqrt{n})$ and $\max_{1\le h\le s_n} |T_d(h)| = O_{\mathbb{P}}(\sqrt{s_n})$ by (SA-17). For the proxy, its three-term expression from the preceding proof is bounded above by

$$\frac{A_k^2}{\xi_d k} + \frac{(A_n - A_k)^2}{\xi_d(n-k)}.$$

Apply (SA-23) to the i.i.d. centered increments $\mathbb{1}(d_i = d)\varepsilon_i(d)$ to bound the first maximum by $O_{\mathbb{P}}(\log\log s_n)$. For the second, $n - k \ge n - s_n \sim n$ and the maximal inequality (SA-17) give $O_{\mathbb{P}}(1)$. The empirical and proxy criteria are also nonnegative. Summing their upper bounds and then over the two arms therefore gives the required left-endpoint order. Reversing the observation ordering likewise gives the right-endpoint process, with counts and residual sums reversed. A finite union over p completes the proof.

SA-5.27. Proof of Lemma SA-26. Suppress the fixed ordering. Let $N_d(k) = \sum_{i=1}^{k} \mathbb{1}(d_i = d)$, and $n_d = N_d(n)$. Writing F_d for the law of $\varepsilon_i(d)$, Assumption SA-2 gives

$$\mathcal{L}\big((\varepsilon_i(d_i))_{i=1}^n \,\big|\, \mathbf{X}, \mathbf{d}\big) = \bigotimes_{i=1}^{n} F_{d_i}.$$

Complete the two observed arm streams with independent draws from F_d, giving $(E_{d,h})_{h\ge1}$. The observed entries belong to disjoint individuals, so

$$\mathcal{L}\big((E_{0,h})_{h\ge1}, (E_{1,h})_{h\ge1} \,\big|\, \mathbf{X}, \mathbf{d}\big) = F_0^{\otimes\mathbb{N}} \otimes F_1^{\otimes\mathbb{N}}.$$

The two i.i.d. streams are therefore mutually independent and independent of assignments. Only one error per individual is used, so latent potential errors may be dependent.

Let $T_d(h)$ be the partial sums of the stream standardized by σ_d. The scalar strong approximation [2, Theorem 2.6.2, p. 107], applied independently to the two streams, then supplies independent Brownian motions W_d, independent of the assignments, with

$$\max_{1\le h\le n} |T_d(h) - W_d(h)| = O_{\mathbb{P}}(\log n).$$

Let

$$\mathcal{C}_n = \bigcap_{d=0}^{1} \{n_d \ge \xi_d n/2\}, \qquad \mathbb{P}(\mathcal{C}_n^c) = O(e^{-cn})$$

by (SA-19). Put $t = k/n$. For $n_d > 0$, define

$$u_d = N_d(k)/n_d, \qquad B_d(u) = \frac{W_d(n_d u) - uW_d(n_d)}{\sqrt{n_d}}.$$

For $n_d = 0$, take an additional independent standard bridge. Conditional on assignments, (B_0, B_1) always has the product standard-bridge law. Thus the bridges are mutually independent and independent of assignments. Work on $\mathcal{C}_n$ below.

The definition of the proxy means gives the exact normalization

$$Z_{k,\ell,d} = \frac{T_d(N_d(k)) - tT_d(n_d)}{\sqrt{\xi_d n t(1-t)}}.$$

Since

$$T_d(N_d(k)) - tT_d(n_d) = \{T_d(N_d(k)) - u_d T_d(n_d)\} + (u_d - t)T_d(n_d),$$

the KMT approximation gives, uniformly for $r \le k \le n - r$,

$$Z_{k,\ell,d} = \sqrt{\frac{n_d}{n\xi_d}} \frac{B_d(u_d)}{\sqrt{t(1-t)}} + \frac{(u_d - t)T_d(n_d)}{\sqrt{\xi_d n t(1-t)}} + O_{\mathbb{P}}\left(\frac{\log n}{\sqrt{r}}\right).$$

Conditional on n_d, the assignments in this ordering are a uniform permutation of n_d ones and $n - n_d$ zeros. For a given cut, expose the smaller side, with size $h = \min\{k, n-k\}$. If $V_1, \dots, V_h$ are the exposed indicators, the Doob martingale for their total has, at step j,

$$M_j = \sum_{i=1}^{j} V_i + \frac{h-j}{n-j}\left(n_d - \sum_{i=1}^{j} V_i\right), \qquad 0 \le j \le h.$$

For $\mathcal{F}_j = \sigma(V_1, \dots, V_j)$, let

$$p_j = \mathbb{E}[V_j \mid \mathcal{F}_{j-1}, n_d] = \frac{n_d - \sum_{i=1}^{j-1} V_i}{n - j + 1}.$$

Subtraction of the two martingale formulas then gives

$$M_j - M_{j-1} = \frac{n-h}{n-j}(V_j - p_j), \qquad |M_j - M_{j-1}| \le 1.$$

For $|x| \le 1$, convexity gives $e^{\lambda x} \le (1+x)e^{\lambda}/2 + (1-x)e^{-\lambda}/2$. Since the increment is conditionally centered,

$$\mathbb{E}\left[e^{\lambda(M_j - M_{j-1})} \,\middle|\, \mathcal{F}_{j-1}, n_d \right] \le \cosh\lambda \le e^{\lambda^2/2}, \qquad \lambda \in \mathbb{R}.$$

Iterating this bound and optimizing the exponential Markov inequality gives

$$\mathbb{P}\big(|M_h - M_0| > x \mid n_d \big) \le 2\exp\{-x^2/(2h)\}.$$

Since $|M_h - M_0| = |N_d(k) - (k/n)n_d|$ and $k(1-k/n) \ge h/2$, the union bound gives

$$\mathbb{P}\left(\max_{1 \le k < n} \frac{|N_d(k) - (k/n)n_d|}{\sqrt{k(1-k/n)}} > C\sqrt{\log n} \,\middle|\, n_d \right) \le 2n^{1 - C^2/4}.$$

For a sufficiently large fixed C, define

$$\mathcal{C}_n' = \mathcal{C}_n \cap \bigcap_{\substack{d \in \{0,1\} \\ 1 \le k < n}} \left\{ |N_d(k) - (k/n)n_d| \le C\sqrt{k(1-k/n)\log n} \right\}.$$

Then $\mathbb{P}(\mathcal{C}_n') \to 1$ and, on this event,

$$|u_d - t| \le C\sqrt{\frac{t(1-t)\log n}{n}}.$$

Conditional on assignments, the bridge increment is centered Gaussian, with

$$\mathbb{V}[B_d(u_d) - B_d(t) \mid \mathbf{d}] = |u_d - t| - (u_d - t)^2 \le |u_d - t|,$$
$$\mathbb{V}\left[\frac{B_d(u_d) - B_d(t)}{\sqrt{t(1-t)}} \,\middle|\, \mathbf{d} \right] \le C\sqrt{\frac{\log n}{nt(1-t)}} \le C\sqrt{\frac{\log n}{r}}.$$

A conditional Gaussian tail bound and a union bound over the retained grid together give

$$\max_{r \le k \le n-r} \frac{|B_d(u_d) - B_d(t)|}{\sqrt{t(1-t)}} = O_{\mathbb{P}}\left(\frac{(\log n)^{3/4}}{r^{1/4}} \right).$$

The remaining terms in turn satisfy

$$T_d(n_d) = O_{\mathbb{P}}(\sqrt{n}), \qquad \sqrt{\frac{n_d}{n\xi_d}} - 1 = O_{\mathbb{P}}(n^{-1/2}),$$
$$\max_{1 \le k < n} \frac{|u_d - t|\,|T_d(n_d)|}{\sqrt{\xi_d nt(1-t)}} = O_{\mathbb{P}}(\sqrt{\log n / n}), \qquad \max_{1 \le k < n} \frac{|B_d(t)|}{\sqrt{t(1-t)}} = O_{\mathbb{P}}(\sqrt{\log n}).$$

The first bound follows from the arm-stream maximal inequality, the second from the binomial variance, and the last from Gaussian tails. Together these prove (SA-14).

At $r_n = \lceil (\log n)^8 \rceil$, write $Y_k = (B_0(k/n), B_1(k/n))^\top / \sqrt{(k/n)(1-k/n)}$. Then

$$\max_{r_n \le k \le n - r_n} \|Z_{k,\ell} - Y_k\| = O_{\mathbb{P}}\left((\log n)^{-3} + (\log n)^{-5/4} + \sqrt{\frac{\log n}{n}} \right),$$
$$\max_{r_n \le k \le n - r_n} (\|Z_{k,\ell}\| + \|Y_k\|) = O_{\mathbb{P}}(\sqrt{\log n}).$$

For fixed symmetric $\mathbf{C}$, $|z^\top \mathbf{C} z - y^\top \mathbf{C} y| \le \|\mathbf{C}\|\|z - y\|(\|z\| + \|y\|)$ therefore gives

$$\max_{r_n \le k \le n - r_n} |Z_{k,\ell}^\top \mathbf{C} Z_{k,\ell} - Y_k^\top \mathbf{C} Y_k| = O_{\mathbb{P}}\left((\log n)^{-3/4} + \frac{\log n}{\sqrt{n}} \right) = o_{\mathbb{P}}((\log n)^{-1/2}).$$

SA-5.28. Proof of Theorem SA-27. Put $r_n = \lceil (\log n)^8 \rceil$, $\sigma_d^2 = \mathbb{V}[\varepsilon_i(d)]$, and $\Lambda = (\sigma_0^2, \sigma_1^2)$. The centered i.i.d. vector

$$V_i = \left(\frac{(1-d_i)\varepsilon_i(0)}{\sqrt{1-\xi}}, \frac{d_i \varepsilon_i(1)}{\sqrt{\xi}} \right)^\top$$

is independent of all covariate orderings, and has covariance $\mathrm{diag}(\sigma_0^2, \sigma_1^2)$. The off-diagonal covariance vanishes since the two coordinates are mutually exclusive and each has mean zero. Independence of the latent potential errors is, again, unnecessary. Moreover,

$$\left\| \frac{\sum_{i=1}^k V_{\pi_\ell(i)} - (k/n)\sum_{i=1}^n V_{\pi_\ell(i)}}{\sqrt{k(1-k/n)}} \right\|^2 = \mathscr{I}^{\mathrm{prox}}(k,\ell).$$

This follows by subtracting the two proxy arm means and multiplying by $\sqrt{\xi_d k(n-k)/n}$ in each coordinate. Lemma SA-24 then gives a uniform $o_{\mathbb{P}}(1)$ error between this proxy and the actual retained scores. Define

$$\mathcal{V}_n^* = \bigcap_{\substack{\ell \in [p] \\ d \in \{0,1\}}} \{ n_{\mathrm{L},d}(r_n, \ell) > 0,\ n_{\mathrm{R},d}(n - r_n, \ell) > 0 \}.$$

On this event, the first and last r_n observations in every ordering each contain both arms. Every retained cut is therefore valid. A union bound over the two boundary blocks and p orderings gives

$$\mathbb{P}((\mathcal{V}_n^*)^c) \le 2p\{\xi^{r_n} + (1-\xi)^{r_n}\} = o(1).$$

Lemma SA-26, with the coordinates multiplied by σ_0, σ_1, also gives (SA-21). Under logit time the Gaussian score is

$$Q_\Lambda(t) = \sigma_0^2 U_0(t)^2 + \sigma_1^2 U_1(t)^2,$$

with independent standard O-U components. Its thresholds and intensity are given by Lemma SA-41. The actual endpoint bound is

$$\max_{\substack{(k,j) \in \mathcal{V}_{\mathrm{SSE}} \\ k < r_n \text{ or } k > n - r_n}} \mathscr{I}^{\mathrm{SSE}}(k,j) = O_{\mathbb{P}}(\log\log r_n) = o_{\mathbb{P}}(\log\log n)$$

by Lemma SA-25. Thus Lemma SA-39 applies. The target, its complement in coordinate ℓ, and the other full coordinate ranges have coefficients

$$b - a, \qquad 2 - b + a, \qquad 2, \ldots, 2,$$

whose sum is $2p$. Therefore, for every measurable tie rule,

$$\mathbb{P}(n^a \le \hat{\imath}_{\mathrm{SSE}} \le n^b,\ \hat{\jmath}_{\mathrm{SSE}} = \ell) \to \frac{b-a}{2p},$$
$$\mathbb{P}(n^a \le n - \hat{\imath}_{\mathrm{SSE}} \le n^b,\ \hat{\jmath}_{\mathrm{SSE}} = \ell) \to \frac{b-a}{2p}.$$

The conditioning argument in the proof of Theorem SA-1 then gives $\mathbb{P}(\hat{\jmath}_{\mathrm{SSE}} = \ell) \to 1/p$ and the conditional window limits. Adding the two conditional limits gives

$$\mathbb{P}\left(a \le \frac{\log \min\{\hat{\imath}_{\mathrm{SSE}}, n - \hat{\imath}_{\mathrm{SSE}}\}}{\log n} \le b \,\middle|\, \hat{\jmath}_{\mathrm{SSE}} = \ell \right) \to b - a.$$

The statistic lies in $[0,1]$ under the auxiliary no-split convention. Let $a \downarrow 0$ and $b \uparrow 1$ to exclude endpoint mass and obtain the conditional uniform limit. Summing the joint limits over coordinates in turn gives the corresponding unconditional limits.

For every fixed $u \in \mathbb{R}$, the same joint limit and the uniform proxy error imply, for any deterministic $v_n \downarrow 0$ with $r_n^{-1/2} = o(v_n)$,

$$\liminf_{n\to\infty} \mathbb{P}\left(\begin{array}{c} n^a \le \hat{\imath}_{\mathrm{SSE}} \le n^b,\ \hat{\jmath}_{\mathrm{SSE}} = \ell, \\ q_\Lambda(Z_{\hat{\imath}_{\mathrm{SSE}},\ell}) \ge w_n(u;\Lambda) - v_n \end{array} \right) \ge \exp\{-(2p-b+a)\kappa_\Lambda e^{-u}\} - \exp\{-2p\kappa_\Lambda e^{-u}\}, \tag{SA-73}$$

where $q_\Lambda(\mathbf{z}) = \sigma_0^2 z_0^2 + \sigma_1^2 z_1^2$ for $\mathbf{z} = (z_0, z_1)^\top \in \mathbb{R}^2$. To justify this lower bound, define the target and complementary valid sets and their actual maxima by

$$\mathcal{T}_n = \{(k,j) \in \mathcal{V}_{\mathrm{SSE}} : j = \ell,\ n^a \le k \le n^b\}, \qquad \mathcal{C}_n = \mathcal{V}_{\mathrm{SSE}} \setminus \mathcal{T}_n,$$
$$M_{T,n} = \max_{(k,j)\in\mathcal{T}_n} \mathscr{I}^{\mathrm{SSE}}(k,j), \qquad M_{C,n} = \max_{(k,j)\in\mathcal{C}_n} \mathscr{I}^{\mathrm{SSE}}(k,j).$$

Use $-\infty$ for empty maxima. Let $\mathcal{E}_n$ denote the empirical event in (SA-73) and put

$$\delta_n = \max_{\substack{(k,j)\in\mathcal{V}_{\mathtt{SSE}}\\ r_n\le k\le n-r_n}} |\mathscr{I}^{\mathtt{SSE}}(k,j) - q_\Lambda(Z_{k,j})| = O_{\mathbb{P}}(r_n^{-1/2}).$$

Since $r_n^{-1/2} = o(v_n)$,

$$\{M_{C,n} < w_n(u;\Lambda) \le M_{T,n}\} \cap \{\delta_n \le v_n\} \subseteq \mathcal{E}_n,$$
$$\mathbb{P}(\mathcal{E}_n) \ge \mathbb{P}(M_{C,n} < w_n(u;\Lambda)) - \mathbb{P}(M_{C,n} < w_n(u;\Lambda),\ M_{T,n} < w_n(u;\Lambda)) - o(1).$$

The continuous joint limit then gives (SA-73).

SA-5.29. Proof of Lemma SA-28. Use $Z_{k,\ell}$, Λ, and r_n from the preceding proofs, and set

$$v_\Lambda = \left(-\sigma_0(1-\xi)^{-1/2},\ \sigma_1\xi^{-1/2}\right)^\top.$$

Both coordinates of v_Λ are nonzero. For a left child $S_{k,\ell} = \{\mathbf{x}\in\mathcal{X} : x_\ell \le x_{\pi_\ell(k),\ell}\}$ with $n^a \le k \le n^b$, Lemma SA-40, applied with cutoff n^a, gives

$$\max_{\substack{\ell\in[p],\ n^a\le k\le n^b\\ (k,\ell)\in\mathcal{V}_{\mathtt{SSE}}}} \sqrt{k}\,|\Delta(S_{k,\ell}) - \bar\mu_{\mathtt{L},1}(k,\ell) + \bar\mu_{\mathtt{L},0}(k,\ell)| = O_{\mathbb{P}}(n^{-a/2}).$$

This uses the left-child calculation in the proof of Lemma SA-16 on $\mathcal{A}_n(n^a)$ in (SA-67), whose complement has probability $O(n^{-a})$. On $\mathcal{V}_n^*$, every cut in this window is valid eventually, and $\mathbb{P}(\mathcal{V}_n^*)\to 1$. For this calculation, put

$$P_{\mathtt{L}} = \bar\mu_{\mathtt{L},1}(k,\ell) - \bar\mu_{\mathtt{L},0}(k,\ell), \qquad P_{\mathtt{R}} = \bar\mu_{\mathtt{R},1}(k,\ell) - \bar\mu_{\mathtt{R},0}(k,\ell).$$

The definitions, maximal inequality, and Lemmas SA-26 and SA-35 give, uniformly on the window,

$$P_{\mathtt{R}} = O_{\mathbb{P}}(n^{-1/2}), \qquad \|Z_{k,\ell}\| = O_{\mathbb{P}}(\sqrt{\log\log n}), \qquad v_\Lambda^\top Z_{k,\ell} = \sqrt{k(1-k/n)}(P_{\mathtt{L}} - P_{\mathtt{R}}).$$

Hence $\sqrt{k}P_{\mathtt{L}} = O_{\mathbb{P}}(\sqrt{\log\log n})$ and

$$\sqrt{k}\,\Delta(S_{k,\ell}) - v_\Lambda^\top Z_{k,\ell} = \sqrt{k}\{\Delta(S_{k,\ell}) - P_{\mathtt{L}}\} + \{1-\sqrt{1-k/n}\}\sqrt{k}P_{\mathtt{L}} + \sqrt{k(1-k/n)}P_{\mathtt{R}}.$$

Since $1-\sqrt{1-k/n} \le k/n$,

$$\max_{\substack{\ell\in[p],\ n^a\le k\le n^b\\ (k,\ell)\in\mathcal{V}_{\mathtt{SSE}}}} \left|\sqrt{k}\,\Delta(S_{k,\ell}) - v_\Lambda^\top Z_{k,\ell}\right| = O_{\mathbb{P}}\left(n^{-a/2} + n^{b-1}\sqrt{\log\log n} + n^{(b-1)/2}\right) = o_{\mathbb{P}}(1). \tag{SA-74}$$

Reversing the ordering similarly gives the same result for a small right child, with the sign of the projection reversed.

Take $v_n = (\log n)^{-1/2}$ and choose the fixed u so that $2p\kappa_\Lambda e^{-u} = 1$. Equation (SA-73) then gives a liminf of at least

$$e^{-1}\left\{e^{(b-a)/(2p)} - 1\right\} \ge \frac{b-a}{2pe}$$

for a selected target split whose proxy score is at least $w_n(u;\Lambda) - v_n$.

First suppose $\sigma_0^2 = \sigma_1^2$. For fixed $0 < \gamma \le 1/4$, define

$$\mathcal{B}_n(\gamma) = \bigcup_{n^a\le k\le n^b} \left\{ \begin{array}{c} q_\Lambda(Z_{k,\ell}) \ge w_n(u;\Lambda) - v_n, \\ |v_\Lambda^\top Z_{k,\ell}| < \gamma\|v_\Lambda\|\|Z_{k,\ell}\| \end{array} \right\}.$$

The proxy vectors are defined at every cut, so no validity restriction is needed in this upper-bound event. In the coupling of Lemma SA-26, put

$$Y_k = \frac{(B_0(k/n), B_1(k/n))^\top}{\sqrt{(k/n)(1-k/n)}}, \qquad e_n = (\log n)^{-1/4},$$

and define

$$\mathcal{K}_n = \left\{\max_{r_n\le k\le n-r_n} |q_\Lambda(Z_{k,\ell}) - q_\Lambda(Y_k)| \le v_n\right\} \cap \left\{\max_{r_n\le k\le n-r_n} \|Z_{k,\ell} - Y_k\| \le e_n\right\}.$$

By the quantitative bounds in that lemma, $\mathbb{P}(\mathcal{K}_n) \to 1$. On $\mathcal{B}_n(\gamma) \cap \mathcal{K}_n$, any candidate satisfying the radial and angular conditions obeys $\|Y_k\| \geq R_n$, where

$$R_n^2 = \frac{w_n(u;\Lambda) - 2v_n}{\sigma_0^2} = 2\{\log\log n + \log\log\log n + u\} + o(1).$$

Moreover,

$$\frac{|v_\Lambda^\top Y_k|}{\|v_\Lambda\|\|Y_k\|} \leq \frac{\gamma\|Z_{k,\ell}\| + e_n}{\|Y_k\|} \leq \gamma + \frac{(1+\gamma)e_n}{R_n} \leq 2\gamma$$

for all sufficiently large n. The two angular arcs

$$\left\{\mathbf{e} \in \mathbb{R}^2 : \|\mathbf{e}\| = 1, \frac{|v_\Lambda^\top \mathbf{e}|}{\|v_\Lambda\|} \leq 2\gamma\right\}$$

have total length at most $C\gamma$. Choose a net $\mathcal{N}_n$ with angular mesh at most $1/R_n$. Then

$$|\mathcal{N}_n| \leq C(\gamma R_n + 1), \qquad \max_{\mathbf{e}\in\mathcal{N}_n} \mathbf{e}^\top Y_k \geq R_n \cos(1/R_n)$$

whenever the radial and angular conditions hold.

Each such projection, under logit time, is a scalar standard O-U process on an interval of length $T_n = (b-a)\log n + o(1)$. For $x \geq 1$,

$$\mathbb{P}\left(\sup_{0\leq t\leq T} U(t) \geq x\right) \leq C(T+1)xe^{-x^2/2}, \qquad T \geq 0.$$

To verify the bound, write $U(t) = e^{-t/2}\{G + W(e^t - 1)\}$, where $G \sim \mathsf{N}(0,1)$ is independent of W. On an interval of length $h = x^{-2}$, a crossing of x implies $\sup_{0\leq s\leq e^h-1}(G + W(s)) \geq x$. Put $v = e^h - 1$. The reflection formula in [2, Theorem 1.5.1, Eq. (1.5.1), p. 43] then gives, conditional on $G = g$,

$$\mathbb{P}\left(\sup_{0\leq s\leq v} W(s) \geq x - g\right) = \begin{cases} 2\{1 - \Phi((x-g)/\sqrt{v})\}, & g < x, \\ 1, & g \geq x, \end{cases} \leq 2\{1 - \Phi((x-g)/\sqrt{v})\}.$$

Since $G + W(v) \sim \mathsf{N}(0, 1+v)$, integration over G gives

$$\mathbb{P}\left(\sup_{0\leq t\leq h} U(t) \geq x\right) \leq 2\{1 - \Phi(xe^{-h/2})\} \leq Cx^{-1}e^{-x^2/2}.$$

Here Φ is the standard normal distribution function. The Gaussian tail bound and $x^2(1-e^{-h}) \leq 1$ give the last inequality. By stationarity and a cover of $[0,T]$ by at most $Tx^2 + 1$ such intervals,

$$\mathbb{P}\left(\sup_{0\leq t\leq T} U(t) \geq x\right) \leq C(Tx^2+1)x^{-1}e^{-x^2/2} \leq C(T+1)xe^{-x^2/2}.$$

Apply this bound to the angular net, with $x = R_n\cos(1/R_n)$. Since $x^2 = R_n^2 - 1 + o(1)$,

$$\mathbb{P}(\mathcal{B}_n(\gamma)) \leq \mathbb{P}(\mathcal{K}_n^c) + C(T_n+1)(\gamma R_n^2 + R_n)e^{-R_n^2/2},$$
$$e^{-R_n^2/2} = \frac{e^{-u+o(1)}}{(\log n)(\log\log n)}, \qquad R_n^2 \sim 2\log\log n,$$
$$\limsup_{n\to\infty} \mathbb{P}(\mathcal{B}_n(\gamma)) \leq C\gamma(b-a)e^{-u} = \frac{C\gamma(b-a)}{2p}.$$

Here $\kappa_\Lambda = 1$, so $e^{-u} = 1/(2p)$. Choose $\gamma > 0$ small enough that $C\gamma \leq 1/(2e)$.

Now suppose $\sigma_1^2 > \sigma_0^2$, interchanging the arms if necessary. Choose $\rho > 0$ with $|v_{\Lambda,0}|\rho \leq |v_{\Lambda,1}|/2$ and define

$$\mathcal{B}_n^{\text{cone}} = \bigcup_{n^a\leq k\leq n^b} \left\{\begin{array}{c} q_\Lambda(Z_{k,\ell}) \geq w_n(u;\Lambda) - v_n, \\ |Z_{k,\ell,0}| > \rho|Z_{k,\ell,1}| \end{array}\right\}.$$

On $|z_0| > \rho|z_1|$,

$$q_\Lambda(\mathbf{z}) \leq \lambda_{\text{bad}}\|\mathbf{z}\|^2, \qquad \lambda_{\text{bad}} = \frac{\sigma_1^2 + \sigma_0^2\rho^2}{1+\rho^2} < \sigma_1^2.$$

But the coupling and Lemma SA-35 with $d = 2$ give

$$\max_{r_n \le k \le n - r_n} \|Z_{k,\ell}\|^2 = 2 \log\log n \{1 + o_{\mathbb{P}}(1)\}.$$

Thus

$$\mathbb{P}(\mathcal{B}_n^{\mathrm{cone}}) \le \mathbb{P}\left(\max_{r_n \le k \le n - r_n} \|Z_{k,\ell}\|^2 \ge \frac{w_n(u;\Lambda) - v_n}{\lambda_{\mathrm{bad}}} \right) \to 0,$$

$$|z_0| \le \rho |z_1| \quad \Longrightarrow \quad |v_\Lambda^\top \mathbf{z}| \ge (|v_{\Lambda,1}| - |v_{\Lambda,0}|\rho)|z_1| \ge \frac{|v_{\Lambda,1}|}{2\sqrt{1+\rho^2}} \|\mathbf{z}\|.$$

Let $\mathcal{E}_n$ be the selected event in (SA-73), with $v_n = (\log n)^{-1/2}$, and let

$$\mathcal{B}_n = \begin{cases} \mathcal{B}_n(\gamma), & \sigma_0^2 = \sigma_1^2, \\ \mathcal{B}_n^{\mathrm{cone}}, & \sigma_1^2 > \sigma_0^2. \end{cases}$$

Interchange arms when $\sigma_0^2 > \sigma_1^2$. In both cases,

$$\begin{aligned} \liminf_{n\to\infty} \mathbb{P}(\mathcal{E}_n \setminus \mathcal{B}_n) &\ge \frac{b-a}{2pe} - \frac{b-a}{4pe} = \frac{b-a}{4pe}, \\ q_\Lambda(Z_{i_{\mathtt{SSE}},\ell}) &\ge w_n(u;\Lambda) - v_n = 2\max\{\sigma_0^2, \sigma_1^2\} \log\log n \{1 + o(1)\}, \\ |v_\Lambda^\top Z_{i_{\mathtt{SSE}},\ell}| &\ge c \|Z_{i_{\mathtt{SSE}},\ell}\| \ge c\sqrt{\log\log n} \qquad \text{on } \mathcal{E}_n \setminus \mathcal{B}_n \end{aligned}$$

for a fixed $c > 0$. Since (SA-74) is uniform, it holds at the selected rank. For a sufficiently small fixed $C_1 > 0$, with an additional failure probability $o(1)$,

$$n(S_{i_{\mathtt{SSE}},\ell}) \Delta(S_{i_{\mathtt{SSE}},\ell})^2 \ge C_1^2 \log\log n, \qquad n(S_{i_{\mathtt{SSE}},\ell}) \le n^b.$$

The constants do not depend on a, b, and $C_2 = (4pe)^{-1}$ is admissible. Reversal similarly gives the right-boundary conclusion.

SA-5.30. Proof of Theorem SA-29. Fix $0 < a < b < 1$. Take the minima of the constants in Lemma SA-28 over the $2p$ windows and denote them by $C_1, C_2 > 0$. For large n, the events that the selected rank and coordinate lie in one of the $2p$ reflected windows are disjoint. On each event supplied by Lemma SA-28, the corresponding child $\mathtt{t}_S$ has size at most n^b and

$$|\hat{\tau}_{\mathtt{SSE}}(\mathtt{t}_S) - \tau| = |\Delta(\mathtt{t}_S)| \ge C_1 n^{-b/2} \sqrt{\log\log n}.$$

The SSE rule uses the DIM contrast for terminal estimation. The supremum error is at least the selected child's error, so the disjoint windows give

$$\liminf_{n\to\infty} \mathbb{P}\left(\sup_{\mathbf{x} \in \mathcal{X}} |\hat{\tau}_{\mathtt{SSE}}(\mathbf{x}) - \tau(\mathbf{x})| \ge C_1 n^{-b/2} \sqrt{\log\log n} \right) \ge 2pC_2(b-a).$$

The constants do not depend on a, b. Letting $a \downarrow 0$ therefore gives a lower bound $2pC_2 b \ge C_2 b$, with the same constants used below.

For the pointwise assertion, take the event for a fixed coordinate ℓ and its left window. Its threshold is at least the $\lceil n^a \rceil$th order statistic. The beta-moment calculation in the proof of Theorem SA-19 then shows that this threshold exceeds $\eta_n n^{a-1}$ except on an event of probability $o(1)$. Hence the entire strip $x_\ell \le \eta_n n^{a-1}$ belongs to the selected child and has its same fitted value and error bound. Reverse the ordering for the right strip. Every point of $\mathcal{X}_n(a, \eta_n)$ belongs to at least one of these fixed $2p$ strips. Taking the smallest of their common probability constants therefore proves the uniform infimum statement. The high-probability order-statistic event is intersected with the selected-score event by a union bound, so no independence between the two is required.

SA-5.31. Proof of Theorem SA-30. Fix $0 < a < b < 1$ with $2^K \log n = o(n^{a/2}\sqrt{\log\log n})$. The only additional step relative to the DIM proof is to obtain the same type of root child event for the tree grown by the `SSE` rule. On $\mathcal{V}_{\mathtt{SSE}} \neq \varnothing$, let $\mathtt{t}_n$ be the smaller selected root child, choosing the left child in a tie, and put $M_n = n(\mathtt{t}_n)$. Define

$$\mathcal{E}_{\mathtt{SSE},n} = \left\{ \mathcal{V}_{\mathtt{SSE}} \neq \varnothing, \quad n^a \le M_n \le n^b, \quad \sqrt{M_n} |\Delta(\mathtt{t}_n)| \ge c_{\mathrm{root}} \sqrt{\log\log n} \right\}.$$

The selected-child conditions are imposed only when the root is split. Lemma SA-28, summed over the two boundary sides and over coordinates, gives constants $c_{\mathrm{root}}, q_{\mathrm{root}} > 0$ with

$$\liminf_{n\to\infty} \mathbb{P}(\mathcal{E}_{\mathtt{SSE},n}) \ge q_{\mathrm{root}}(b-a).$$

Take $c_{\mathrm{root}} = C_1$ and $q_{\mathrm{root}} = 2pC_2$ from Lemma SA-28. This lemma supplies the scalar contrast bound in $\mathcal{E}_{\mathtt{SSE},n}$.

Let $\mathcal{P}(\mathsf{t}_n)$ denote the terminal descendants of t_n. Valid splitting again preserves both construction arms in each descendant. On $\mathcal{E}_{\mathtt{SSE},n}$,

$$M_n \in [n^a, n^b], \qquad |\mathcal{P}(\mathsf{t}_n)| \le 2^K, \qquad \frac{2^K \log n}{n^{a/2}\sqrt{\log\log n}} \to 0.$$

Applying Corollary SA-50 then gives

$$\mathbb{P}\left(\mathcal{E}_{\mathtt{SSE},n},\ \max_{\mathsf{t}'\in\mathcal{P}(\mathsf{t}_n)} |\Delta(\mathsf{t}')| \ge \frac{c_{\mathrm{root}}}{2} n^{-b/2}\sqrt{\log\log n}\right) \ge \mathbb{P}(\mathcal{E}_{\mathtt{SSE},n}) - o(1).$$

For every terminal descendant t' and every $\mathbf{x} \in \mathsf{t}'$,

$$\hat{\tau}_{\mathtt{SSE}}(\mathbf{x}) - \tau(\mathbf{x}) = \Delta(\mathsf{t}'),$$

because Definition SA-6 computes terminal treatment effects on a tree grown by the `SSE` rule using the `DIM` estimator. Taking the supremum over $\mathbf{x} \in \mathcal{X}$ and taking the liminf therefore gives

$$\liminf_{n\to\infty} \mathbb{P}\left(\sup_{\mathbf{x}\in\mathcal{X}} |\hat{\tau}_{\mathtt{SSE}}(\mathbf{x}) - \tau(\mathbf{x})| \ge \frac{c_{\mathrm{root}}}{2} n^{-b/2}\sqrt{\log\log n}\right) \ge q_{\mathrm{root}}(b-a).$$

Set

$$c_{\mathtt{SSE}} = c_{\mathrm{root}}/2, \qquad q_{\mathtt{SSE}} = q_{\mathrm{root}}/2.$$

Then $q_{\mathrm{root}}(b-a) = 2q_{\mathtt{SSE}}(b-a)$, proving (SA-15). Taking $a = b/2$ gives its stated specialization. For $2\rho_K < b$, (SA-72) verifies the growth condition for every fixed $a \in (2\rho_K, b)$. The error event and constants again do not depend on a, so

$$\liminf_{n\to\infty} \mathbb{P}\left(\sup_{\mathbf{x}\in\mathcal{X}} |\hat{\tau}_{\mathtt{SSE}}(\mathbf{x}) - \tau(\mathbf{x})| \ge c_{\mathtt{SSE}} n^{-b/2}\sqrt{\log\log n}\right) \ge \sup_{a\in(2\rho_K,b)} 2q_{\mathtt{SSE}}(b-a) = 2q_{\mathtt{SSE}}(b-2\rho_K).$$

This proves (SA-16).

SA-5.32. Proof of Theorem SA-31. For any fitted partition, the NSS-SSE terminal estimator is the DIM terminal contrast, with the same zero convention when either treatment arm is absent. The proof of Theorem SA-21 already bounds its integrated error simultaneously over all axis-aligned partitions with at most 2^K leaves. It therefore applies to the outcome-dependent NSS-SSE partition and gives both stated bounds for every fixed $1 \le q < \infty$.

SA-5.33. Proof of Theorem SA-32. Let $M = n_{\mathsf{T}}$, $N = n_\tau$, $a = b/2$, and

$$\mathcal{E}_n^{\mathtt{SSE}} = \{\mathcal{V}_{\mathtt{SSE}}(M) \ne \varnothing, \quad M^a \le \hat{\imath}_{\mathtt{SSE}} \le M^b\}.$$

Theorem SA-27, applied with sample size M and summed over coordinates, gives $\mathbb{P}(\mathcal{E}_n^{\mathtt{SSE}}) \to (b-a)/2 = b/4$. On this event, let $m = \hat{\imath}_{\mathtt{SSE}}$ and let W, R be the population mass and estimation count of the selected left root child. SSE selection, like DIM selection, depends only on the rank permutations, assignments, and outcomes. Equation (SA-30) therefore gives

$$W \mid (\hat{\jmath}_{\mathtt{SSE}}, \hat{\imath}_{\mathtt{SSE}} = m) \sim \mathsf{Beta}(m, M+1-m), \qquad R \mid W \sim \mathsf{Binomial}(N, W).$$

The beta-binomial calculation in the HON-DIM proof, with $c_R = \rho/4$ and $C_R = 2/\rho$, then yields

$$\mathbb{P}\big(R \notin [c_R m, C_R m] \mid \hat{\jmath}_{\mathtt{SSE}},\ \hat{\imath}_{\mathtt{SSE}} = m\big) \le C/m \le CM^{-a}$$

uniformly over the root event.

Order the terminal descendants by tree address and write $N_1, \dots, N_J$ for their estimation counts. Valid splitting again gives $J \le m/2$. The disjoint-pair calculation in the HON-DIM proof similarly gives

$$\mathbb{P}\left(\max_{1\le j\le J} N_j \ge 2 \,\middle|\, \mathcal{D}_{\mathsf{T}},\ R\right) \ge 1 - e^{-\lfloor R/2\rfloor/J} \ge q_{\mathrm{occ}} := 1 - e^{-2c_R/3} = 1 - e^{-\rho/6} > 0$$

on $\mathcal{E}_n^{\mathtt{SSE}} \cap \{c_R m \le R \le C_R m\}$ for large n. Hence

$$\mathcal{H}_n^{\mathtt{SSE}} = \mathcal{E}_n^{\mathtt{SSE}} \cap \{c_R m \le R \le C_R m\} \cap \{\max_{1\le j\le J} N_j \ge 2\},$$

$$\liminf_{n\to\infty} \mathbb{P}(\mathcal{H}_n^{\mathtt{SSE}}) \ge q_{\mathrm{occ}} b/4.$$

On $\mathcal{H}_n^{\mathtt{SSE}}$ select $j_\star = \min\{j : N_j \ge 2\}$. This event and selection are measurable with respect to $\mathcal{F}_n = \sigma(\mathcal{D}_{\mathsf{T}}, \tilde{\mathbf{X}})$, where $\tilde{\mathbf{X}}$ contains the estimation covariates. Thus

$$2 \le N_\star = N_{j_\star} \le C_R n^b, \qquad \mathbb{P}(N_{1,\star} N_{0,\star} > 0 \mid \mathcal{F}_n) = 1 - \xi^{N_\star} - (1-\xi)^{N_\star} \ge 2\xi(1-\xi).$$

The selected assignments remain i.i.d. Bernoulli(ξ). Conditional also on these assignments, the observed residuals retain their independent arm laws. The SSE terminal estimate is the DIM contrast, so Lemma SA-51 gives

$$\mathbb{P}\left(\sup_{\mathbf{x}\in\mathcal{X}}|\check{\tau}_{\texttt{SSE}}(\mathbf{x})-\tau(\mathbf{x})|\geq\frac{c_{\text{ac}}}{\sqrt{C_R}}n^{-b/2}\right)\geq 2\xi(1-\xi)q_{\text{ac}}\mathbb{P}(\mathcal{H}_n^{\texttt{SSE}}),$$
$$\liminf_{n\to\infty}\mathbb{P}\left(\sup_{\mathbf{x}\in\mathcal{X}}|\check{\tau}_{\texttt{SSE}}(\mathbf{x})-\tau(\mathbf{x})|\geq\frac{c_{\text{ac}}}{\sqrt{C_R}}n^{-b/2}\right)\geq\frac{\xi(1-\xi)q_{\text{ac}}q_{\text{occ}}}{2}\,b.$$

With the HON-DIM values of c_{ac}, q_{ac} and C_R, the threshold coefficient is $c_{\text{ac}}/\sqrt{C_R}=\sqrt{\rho(\sigma_0^2+\sigma_1^2)}/2$ and the probability bound equals that in Theorem SA-32. No depth restriction enters the bound $J\leq m/2$.

SA-5.34. Proof of Theorem SA-33. HON-SSE estimates each terminal effect by the DIM contrast and sets it to zero when either estimation arm is absent. Given the SSE construction fold, the partition is fixed and independent of the estimation fold, so the leaf-count calculation in the proof of Theorem SA-23 applies. It accordingly gives the expectation and probability bounds for every fixed $1\leq q<\infty$, using Hölder's inequality for $q<2$, with the HON-DIM constants unchanged.

SA-5.35. Proof of Lemma SA-34. For HON, condition on the construction partition and the independent estimation covariates. No symmetry is needed here.

Let $\tilde{\mathbf{X}}$ and $\tilde{\mathbf{d}}$ denote the estimation-fold covariate matrix and assignment vector. For HON-IPW, condition on the construction partition and all estimation covariates. These variables determine the terminal node containing the fixed evaluation point $\mathbf{x}$ and its estimation count. The IPW pseudo-outcomes in that node remain independent of this conditioning and have mean τ. Therefore

$$\mathbb{E}[\check{\tau}_{\texttt{IPW}}(\mathbf{x})\mid\mathsf{T},\tilde{\mathbf{X}}]=\tau(\mathbf{x})\mathbb{1}(n(\mathtt{t}_{\texttt{HON}}(\mathbf{x}))>0).$$

Averaging then yields $\mathbb{E}[\check{\tau}_{\texttt{IPW}}(\mathbf{x})]=\tau(\mathbf{x})-\tau(\mathbf{x})\mathbb{P}(n(\mathtt{t}_{\texttt{HON}}(\mathbf{x}))=0)$. For HON-DIM and HON-SSE, condition additionally on the estimation assignments. When both arms are present, the residual means have conditional expectation zero and their difference is added to τ. If either arm is absent, the entire estimator equals zero. Hence, for $l\in\{\texttt{DIM},\texttt{SSE}\}$,

$$\mathbb{E}[\check{\tau}_l(\mathbf{x})\mid\mathsf{T},\tilde{\mathbf{X}},\tilde{\mathbf{d}}]=\tau(\mathbf{x})\mathbb{1}(n_0(\mathtt{t}_{\texttt{HON}}(\mathbf{x}))>0,\ n_1(\mathtt{t}_{\texttt{HON}}(\mathbf{x}))>0).$$

Taking expectations then gives the missing-arm bias formula. The independence from sample splitting suffices throughout the HON argument, without symmetry.

For NSS, condition on $(\mathbf{X},\mathbf{d})$. We first consider DIM and SSE under the stated marginal symmetry conditions. Put $Z_i=\varepsilon_i(d_i)$. The observed residuals are independent across individuals and have their symmetric arm marginals, so $(Z_1,\dots,Z_n)\overset{d}{=}-(Z_1,\dots,Z_n)$. At any fixed node, because μ_0 and μ_1 are constant, the split gains maximized under Definition SA-5 can be written as

$$\texttt{DIM}:\quad\frac{n(\mathtt{t}_{\texttt{L}})n(\mathtt{t}_{\texttt{R}})}{n(\mathtt{t})}\Big(\frac{1}{n_1(\mathtt{t}_{\texttt{L}})}\sum_{i:\mathbf{x}_i\in\mathtt{t}_{\texttt{L}}}d_i\varepsilon_i(1)-\frac{1}{n_0(\mathtt{t}_{\texttt{L}})}\sum_{i:\mathbf{x}_i\in\mathtt{t}_{\texttt{L}}}(1-d_i)\varepsilon_i(0)-\frac{1}{n_1(\mathtt{t}_{\texttt{R}})}\sum_{i:\mathbf{x}_i\in\mathtt{t}_{\texttt{R}}}d_i\varepsilon_i(1)+\frac{1}{n_0(\mathtt{t}_{\texttt{R}})}\sum_{i:\mathbf{x}_i\in\mathtt{t}_{\texttt{R}}}(1-d_i)\varepsilon_i(0)\Big)^2,$$

and

$$\begin{aligned}\texttt{SSE}:\quad&\frac{n_1(\mathtt{t}_{\texttt{L}})n_1(\mathtt{t}_{\texttt{R}})}{n_1(\mathtt{t})}\Big(\frac{1}{n_1(\mathtt{t}_{\texttt{L}})}\sum_{i:\mathbf{x}_i\in\mathtt{t}_{\texttt{L}}}d_i\varepsilon_i(1)-\frac{1}{n_1(\mathtt{t}_{\texttt{R}})}\sum_{i:\mathbf{x}_i\in\mathtt{t}_{\texttt{R}}}d_i\varepsilon_i(1)\Big)^2\\&+\frac{n_0(\mathtt{t}_{\texttt{L}})n_0(\mathtt{t}_{\texttt{R}})}{n_0(\mathtt{t})}\Big(\frac{1}{n_0(\mathtt{t}_{\texttt{L}})}\sum_{i:\mathbf{x}_i\in\mathtt{t}_{\texttt{L}}}(1-d_i)\varepsilon_i(0)-\frac{1}{n_0(\mathtt{t}_{\texttt{R}})}\sum_{i:\mathbf{x}_i\in\mathtt{t}_{\texttt{R}}}(1-d_i)\varepsilon_i(0)\Big)^2.\end{aligned}$$

For $l\in\{\texttt{DIM},\texttt{SSE}\}$ and a residual vector $\mathbf{z}=(z_1,\dots,z_n)^\top$, let $\mathcal{V}_l(\mathtt{t};\mathbf{z})$ and $\mathscr{I}_{\mathbf{z}}^l(\mathtt{t};k,j)$ denote the valid candidate set and split score at a fixed node $\mathtt{t}$ when the responses are $\mu_{d_i}+z_i$, with covariates and assignments held fixed. Here k is the cut rank within $\mathtt{t}$ along coordinate j. Write $\mathcal{P}_l(\mathbf{z})$ for the resulting terminal partition. With $\boldsymbol{Z}=(Z_1,\dots,Z_n)^\top$, the sign flip gives

$$\mathcal{V}_l(\mathtt{t};\boldsymbol{Z})=\mathcal{V}_l(\mathtt{t};-\boldsymbol{Z}),\qquad\mathscr{I}_{\boldsymbol{Z}}^l(\mathtt{t};k,j)=\mathscr{I}_{-\boldsymbol{Z}}^l(\mathtt{t};k,j),\qquad(k,j)\in\mathcal{V}_l(\mathtt{t};\boldsymbol{Z}).$$

The prescribed tie rule therefore selects the same children. By the assumed invariance of the final partition after stopping and pruning,

$$\mathcal{P}_l(\boldsymbol{Z})=\mathcal{P}_l(-\boldsymbol{Z}).$$

Let $\mathtt{t}_l$ denote the selected NSS terminal node containing $\mathbf{x}$ for estimator l. For $l\in\{\texttt{DIM},\texttt{SSE}\}$, this node and its treatment counts are unchanged by the sign flip. Let $H_l=\mathbb{1}(n_0(\mathtt{t}_l)>0,\ n_1(\mathtt{t}_l)>0)$ and take the residual contrast W_l as zero when either arm is absent. Then H_l is even, W_l is odd, and $\hat{\tau}_l(\mathbf{x})=\tau(\mathbf{x})H_l+W_l$. These random variables are integrable because $|W_l|\leq 2\max_{1\leq i\leq n}|Z_i|$ and each observed residual has a finite first moment. Conditional symmetry and oddness therefore give

$$\mathbb{E}[W_l\mid\mathbf{X},\mathbf{d}]=\mathbb{E}[W_l(-Z_1,\dots,-Z_n)\mid\mathbf{X},\mathbf{d}]=-\mathbb{E}[W_l\mid\mathbf{X},\mathbf{d}]=0.$$

This is the sign-involution argument in the proof of Lemma SA-8. Taking expectations of $\hat{\tau}_l(\mathbf{x}) = \tau(\mathbf{x})H_l + W_l$ therefore gives

$$\mathbb{E}[\hat{\tau}_l(\mathbf{x})] = \tau(\mathbf{x})\mathbb{P}\big(n_0(\mathtt{t}_l) > 0,\ n_1(\mathtt{t}_l) > 0\big) = \tau(\mathbf{x}) - \tau(\mathbf{x})\mathbb{P}\big(n_0(\mathtt{t}_l) = 0 \text{ or } n_1(\mathtt{t}_l) = 0\big).$$

Only the marginal potential-error laws must be symmetric. The pair of errors for an individual therefore need not be jointly centrally symmetric.

Every valid DIM or SSE split places both treatment arms in each child. Early stopping retains the current node, and pruning joins descendants, with

$$n_d(\mathtt{t}) = n_d(\mathtt{t}_\mathtt{L}) + n_d(\mathtt{t}_\mathtt{R}), \qquad d = 0, 1.$$

Thus every terminal node contains both arms whenever the full sample does. If the full sample lacks one arm, the root has no valid split and the estimator is zero. Consequently,

$$H_l = \mathbb{1}\left(0 < \sum_{i=1}^{n} d_i < n\right), \qquad \mathbb{E}[\hat{\tau}_l(\mathbf{x}) \mid \mathbf{X}, \mathbf{d}] = \tau(\mathbf{x})H_l.$$

Independent treatment assignment then gives

$$\mathbb{P}(H_l = 0) = \mathbb{P}(d_1 = \cdots = d_n = 0) + \mathbb{P}(d_1 = \cdots = d_n = 1) = (1-\xi)^n + \xi^n.$$

Therefore

$$\mathbb{E}[\hat{\tau}_l(\mathbf{x})] = \tau(\mathbf{x})\big\{1 - \xi^n - (1-\xi)^n\big\}, \qquad \mathbb{E}[\hat{\tau}_l(\mathbf{x}) \mid H_l = 1] = \tau(\mathbf{x}),$$

where the conditional expectation is defined for $n \geq 2$.

For IPW, put $\tilde{\varepsilon}_i = y_i(d_i - \xi)/\{\xi(1-\xi)\} - \tau$ and $\tilde{\boldsymbol{\varepsilon}} = (\tilde{\varepsilon}_1, \ldots, \tilde{\varepsilon}_n)^\top$. The assumed conditional symmetry and independence from the covariates give

$$\mathcal{L}\big(\tilde{\boldsymbol{\varepsilon}} \mid \mathbf{X}, \mathbf{d}\big) = \mathcal{L}\big(-\tilde{\boldsymbol{\varepsilon}} \mid \mathbf{X}, \mathbf{d}\big).$$

On every nonempty construction node,

$$\hat{\tau}_{\mathtt{IPW}}(\mathtt{t}) = \tau + \frac{1}{n(\mathtt{t})} \sum_{i:\mathbf{x}_i \in \mathtt{t}} \tilde{\varepsilon}_i.$$

The sign flip reverses each child contrast and leaves its square unchanged. The valid candidates and prescribed tie rule are also unchanged. By the partition condition in the lemma, the final partition is unchanged after stopping and pruning. Put $H_{\mathtt{IPW}} = \mathbb{1}(n(\mathtt{t}_{\mathtt{IPW}}) > 0)$ and let $W_{\mathtt{IPW}}$ be the terminal residual average, set to zero when $H_{\mathtt{IPW}} = 0$. Then

$$H_{\mathtt{IPW}}(-\tilde{\boldsymbol{\varepsilon}}) = H_{\mathtt{IPW}}(\tilde{\boldsymbol{\varepsilon}}), \qquad W_{\mathtt{IPW}}(-\tilde{\boldsymbol{\varepsilon}}) = -W_{\mathtt{IPW}}(\tilde{\boldsymbol{\varepsilon}}), \qquad |W_{\mathtt{IPW}}| \leq \max_{1 \leq i \leq n} |\tilde{\varepsilon}_i|, \qquad \mathbb{E}\left[\max_{1 \leq i \leq n} |\tilde{\varepsilon}_i|\right] < \infty, \qquad \mathbb{E}[W_{\mathtt{IPW}} \mid \mathbf{X}, \mathbf{d}] = 0.$$

Under the alternative IPW conditions, independence from the covariates, i.i.d. sampling, and marginal transformed residual symmetry give

$$\mathcal{L}\big(\tilde{\boldsymbol{\varepsilon}} \mid \mathbf{X}\big) = \mathcal{L}\big(-\tilde{\boldsymbol{\varepsilon}} \mid \mathbf{X}\big).$$

Under this alternative, the final IPW partition is assumed to depend only on $\mathbf{X}$ and the transformed outcomes, including any effects of stopping and pruning. The assumed reflection invariance likewise gives the same evenness and oddness identities for $H_{\mathtt{IPW}}$ and $W_{\mathtt{IPW}}$. The integrability bound then yields

$$\mathbb{E}[W_{\mathtt{IPW}} \mid \mathbf{X}] = -\mathbb{E}[W_{\mathtt{IPW}} \mid \mathbf{X}] = 0.$$

Thus $\mathbb{E}[W_{\mathtt{IPW}}] = 0$ under either set of IPW conditions. Consequently,

$$\mathbb{E}[\hat{\tau}_{\mathtt{IPW}}(\mathbf{x})] = \mathbb{E}[\tau(\mathbf{x})H_{\mathtt{IPW}} + W_{\mathtt{IPW}}] = \tau(\mathbf{x})\mathbb{P}(n(\mathtt{t}_{\mathtt{IPW}}) > 0) = \tau(\mathbf{x}) - \tau(\mathbf{x})\mathbb{P}(n(\mathtt{t}_{\mathtt{IPW}}) = 0).$$

Every valid IPW split has nonempty children. Early stopping retains the current node, and pruning joins descendants. Thus every NSS-IPW terminal node contains a construction observation, so $H_{\mathtt{IPW}} = 1$ and $\mathbb{E}[\hat{\tau}_{\mathtt{IPW}}(\mathbf{x})] = \tau(\mathbf{x})$.

SA-6. Numerical Evidence

This section presents Monte Carlo simulations and an empirical resampling illustration for the six causal tree estimators defined in Section SA-3, with $p \in \{1, 2\}$ covariates. In the figures, X denotes the covariate for $p = 1$, with observed values $x_i = x_{i,1}$, and X_1 and X_2 denote the covariates for $p = 2$, with observed values $\mathbf{x}_i = (x_{i,1}, x_{i,2})^\top$. Replication code and data are available at https://github.com/mdcattaneo/replication-CKY_2027_PNAS.

SA-6.1. Numerical Implementation. All numerical procedures implement Definitions SA-4–SA-6 directly. The IPW procedures, in particular, use the known treatment assignment probability ξ in both tree construction and terminal estimation ($\xi = 1/2$ in the simulations and $\xi = 2/3$ in the JTPA study). At each node, candidate splits are evaluated at the distinct observed values of each covariate in increasing order of the coordinate index $j \in [p]$ and threshold value, with ties broken in favor of the first candidate in this order, up to a relative floating-point tolerance. Trees are then grown recursively until reaching depth K or until no valid split remains, allowing splits with zero impurity gain, and trees of depths $K \in \{1, \dots, 5\}$ are obtained by truncating the depth-5 tree. For the honest estimators, the construction sample $\mathcal{D}_{\mathsf{T}}$ has size $n_{\mathsf{T}} = \lfloor n/2 \rfloor$ and the estimation sample $\mathcal{D}_\tau$ contains the remaining $n_\tau = n - n_{\mathsf{T}}$ observations; the simulations use the first $n/2$ observations for $\mathcal{D}_{\mathsf{T}}$, whereas the empirical resampling study draws $\mathcal{D}_{\mathsf{T}}$ uniformly at random without replacement in each replication and uses the same partition across `HON-IPW`, `HON-DIM`, and `HON-SSE`.

SA-6.2. Simulations. We first illustrate Theorem 1 in the main paper using synthetic designs with $p \in \{1, 2\}$. In each replication, we generate $n = 1{,}000$ independent observations with $\mathbf{x}_i \sim \mathsf{Uniform}([0,1]^p)$, $d_i \sim \mathsf{Bernoulli}(1/2)$, and potential outcomes $y_i(d) = \mu_d^\star + \varepsilon_i(d)$, where $\mu_0^\star = 0$, $\mu_1^\star = 1$, and $\varepsilon_i(0), \varepsilon_i(1) \overset{\text{i.i.d.}}{\sim} \mathsf{N}(0,1)$. The covariates, treatment assignments, and potential outcome errors are mutually independent, so the CATE is constant at $\tau^\star = 1$.

For each dimension and each of the six estimators across tree depths $K \in \{1, \dots, 5\}$, we compute pointwise root mean squared error (RMSE), $\{\mathbb{E}[(\hat{\tau}_l(\mathbf{x}) - \tau(\mathbf{x}))^2]\}^{1/2}$ and $\{\mathbb{E}[(\check{\tau}_l(\mathbf{x}) - \tau(\mathbf{x}))^2]\}^{1/2}$ for $l \in \{\mathtt{IPW}, \mathtt{DIM}, \mathtt{SSE}\}$, over 1,000,000 Monte Carlo replications. Pointwise RMSE is evaluated at 201 equally spaced points on $[0,1]$ for $p = 1$ and on a 51×51 equally spaced grid on $[0,1]^2$ for $p = 2$, with $n_{\mathsf{T}} = n_\tau = 500$ for the honest estimators. Across all evaluation points, depths, and estimators, the Monte Carlo standard error is at most 0.51% of the estimated RMSE.

Figures S1 and S2 report the results for $p = 1$ and $p = 2$. Two patterns are evident across all six estimators. First, for each depth K, pointwise RMSE is lowest in the interior of the covariate support and increases toward the boundary, where greedy splits can produce small terminal cells with few observations available for local averaging (Section SA-1.2). Second, pointwise RMSE increases with tree depth K, as deeper partitioning places a larger share of evaluation points near terminal node boundaries.

SA-6.3. Empirical Resampling. We next consider an empirical resampling design based on the National Job Training Partnership Act (JTPA) Study, a randomized evaluation of Title II-A employment and training services for economically disadvantaged adults and out-of-school youths across 16 local service delivery areas [25, 26]. Applicants were randomly assigned with probability $\xi = 2/3$ to a program group eligible for JTPA services or to a control group excluded from those services for 18 months, with subsequent earnings measured from follow-up surveys and administrative records. We use the adult sample from the public-use file `jtpa.dta` distributed through the Statistical Software Components archive at Boston College (http://fmwww.bc.edu/repec/bocode/j/jtpa.dta), which contains 11,204 individuals aged 22–78. The outcome y_i is total earnings in the 30 months following random assignment, the treatment indicator d_i is the randomized offer of JTPA services, and the covariates are preprogram earnings in the year prior to assignment ($x_{i,1}$) and, in the two-covariate design, age in years at assignment ($x_{i,2}$).

To limit the influence of large mass points in preprogram earnings (rounded to three decimal places), we restrict the data prior to resampling. The one-covariate design restricts preprogram earnings to $[0, 6{,}000]$ dollars, while the two-covariate design restricts preprogram earnings to $[0, 5{,}000]$ dollars and age to at most 50 years. Within each restricted sample, we group observations by their exact value of preprogram earnings and randomly subsample each group of size n_g without replacement, retaining $\min\{\max\{1, \lfloor 0.5 n_g \rfloor\}, 30\}$ observations for $p = 1$ and $\min\{\max\{1, \lfloor 0.2 n_g \rfloor\}, 200\}$ observations for $p = 2$. This yields processed samples of size $n = 2{,}475$ for $p = 1$ and $n = 2{,}132$ for $p = 2$.

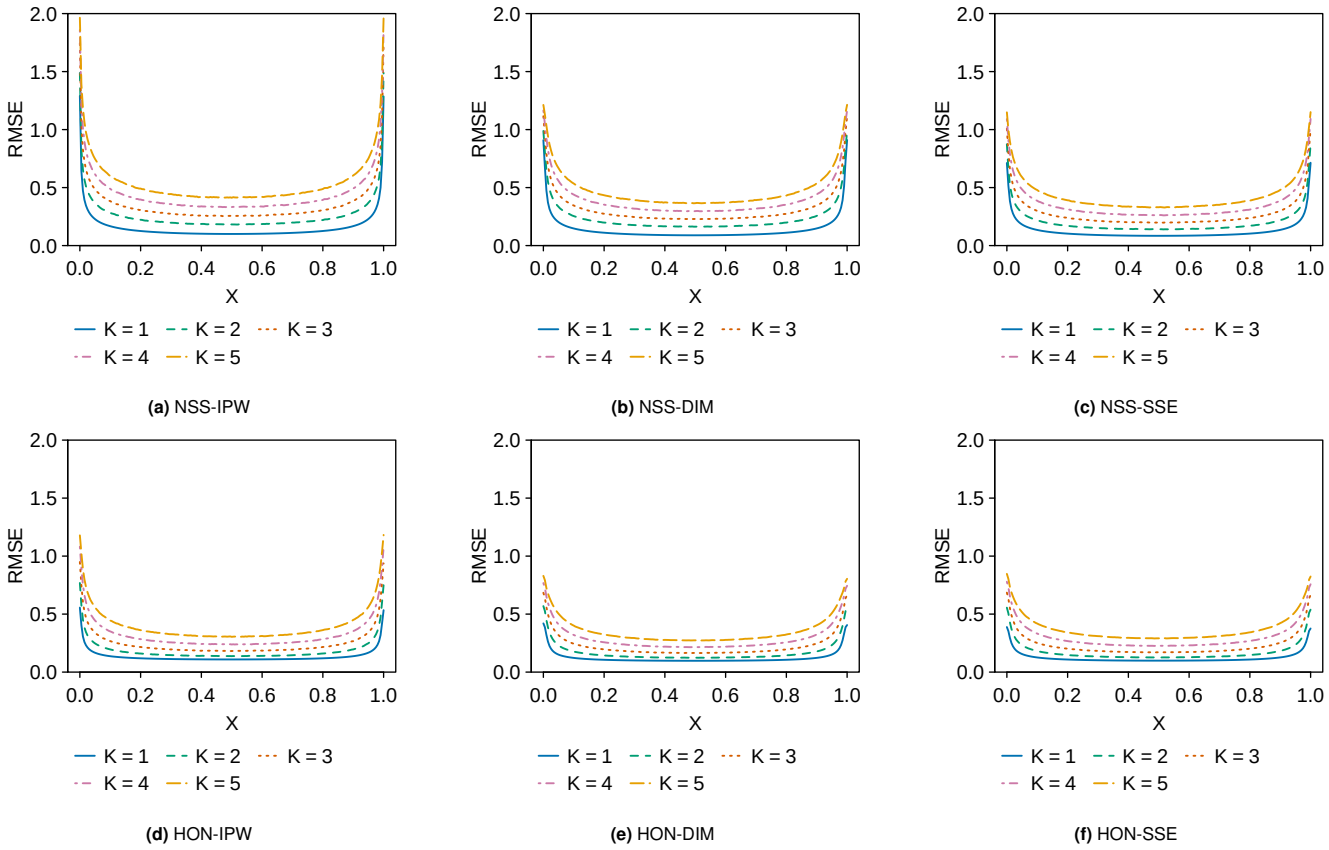
In each of 1,000,000 replications, we hold outcomes and treatment assignments fixed and permute the covariates without replacement within each treatment arm, permuting preprogram earnings in the one-covariate design and both covariates jointly in the two-covariate design. This preserves the empirical within-arm covariate distribution while removing the association between covariates and outcomes within each arm, allowing us to examine estimator behavior when the covariates have mass points and uneven support. Independently in each replication, the honest estimators split the sample at random without replacement into $n_{\mathsf{T}} = \lfloor n/2 \rfloor$ construction observations (1,237 for $p = 1$ and 1,066 for $p = 2$) and $n_\tau = n - n_{\mathsf{T}}$ estimation observations (1,238 for $p = 1$ and 1,066 for $p = 2$).

Because the resampled covariates are discrete and permutations and sample splits are drawn without replacement, this exercise departs from the continuous-covariate, i.i.d. setting of our theoretical results. In addition, pointwise RMSE is evaluated relative to the full-sample difference in treatment-arm outcome means, whereas the IPW estimator uses the design assignment probability $\xi = 2/3$; because empirical covariate distributions can differ across treatment arms, the conditional treatment probability need not equal 2/3 and the IPW transformed outcome need not be centered at the full-sample difference in means.

We then evaluate all six estimators at depths $K \in \{1, \dots, 5\}$ across 201 points chosen at approximately equal index intervals from the ordered distinct preprogram earnings values for $p = 1$, and on a 41×31 equally spaced rectangular grid spanning the marginal 1st–99th percentiles of preprogram earnings and age for $p = 2$, with earnings displayed in thousands of dollars and age in years. Across all evaluation points, depths, and estimators, the Monte Carlo standard error is at most 4.77% ($p = 1$) and 2.58% ($p = 2$) of the estimated RMSE. Figures S3 and S4 report the results. Although the empirical RMSE profiles are more irregular than in the synthetic simulations due to covariate discreteness and uneven support, pointwise RMSE again increases toward the boundaries of the evaluation region across all splitting criteria and both sample-splitting schemes.

References

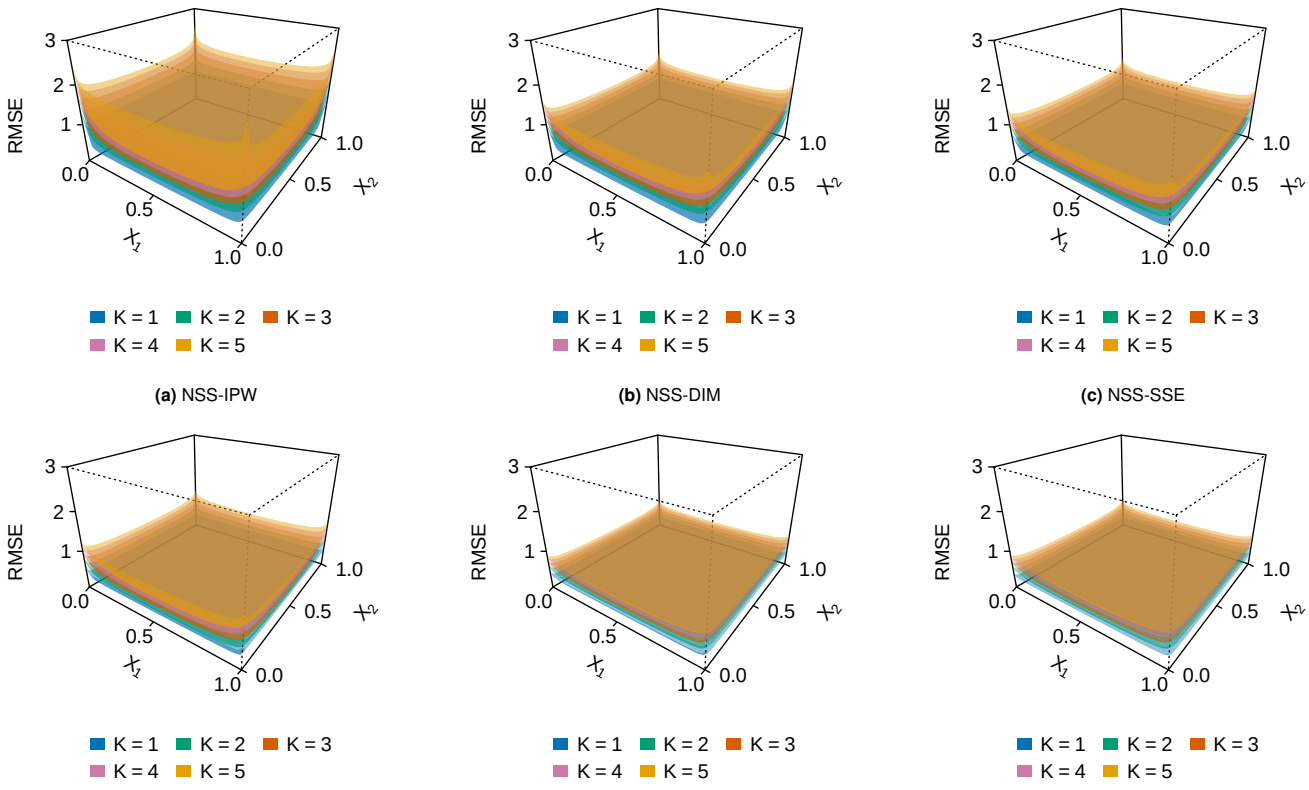
1. F Eicker, The asymptotic distribution of the suprema of the standardized empirical processes. *Annals of Statistics* **7**, 116–138 (1979).
2. M Csörgő, P Révész, *Strong Approximations in Probability and Statistics*, Probability and Mathematical Statistics : a series of monographs and textbooks. (Academic Press), (1981).
3. L Horváth, The maximum likelihood method for testing changes in the parameters of normal observations. *Annals of Statistics* **21**, 671–680 (1993).
4. AI Zhdanov, High excursions of a quadratic form for a Gaussian stationary vector process. *Fundamentalnaya i Prikladnaya Matematika* **23**, 123–144 (2020) In Russian; English translation in Journal of Mathematical Sciences 262(4) (2022), 476–492.
5. A Carbotti, S Cito, DA La Manna, D Pallara, Stability of the Gaussian Faber–Krahn inequality. *Annali di Matematica Pura ed Applicata* **203**, 2185–2198 (2024).
6. S Athey, G Imbens, Y Kong, *causalTree R package documentation*, (2026) GitHub repository, accessed May 2, 2026. https://github.com/susanathey/causalTree.
7. J Xu, T Shinkre, J Brand, *htetree R package documentation*, (2026) Version 0.1.23, accessed May 2, 2026. https://CRAN.R-project.org/package=htetree.
8. H Ishwaran, The effect of splitting on random forests. *Machine Learning* **99**, 75–118 (2015).
9. S Wager, S Athey, Estimation and inference of heterogeneous treatment effects using random forests. *Journal of the American Statistical Association* **113**, 1228–1242 (2018).
10. N Meinshausen, Quantile regression forests. *Journal of Machine Learning Research* **7**, 983–999 (2006).
11. S Athey, J Tibshirani, S Wager, Generalized random forests. *Annals of Statistics* **47**, 1148–1178 (2019).
12. M Behr, Y Wang, X Li, B Yu, Provable Boolean interaction recovery from tree ensemble obtained via random forests. *Proceedings of the National Academy of Sciences* **119**, e2118636119 (2022).
13. MD Cattaneo, JM Klusowski, PM Tian, On the pointwise behavior of recursive partitioning and its implications for heterogeneous causal effect estimation. *Technical report, arXiv preprint arXiv:2211.10805v1* (2022).
14. P Bühlmann, B Yu, Analyzing bagging. *Annals of Statistics* **30**, 927–961 (2002).
15. M Banerjee, IW McKeague, Confidence sets for split points in decision trees. *Annals of Statistics* **35**, 543–574 (2007).
16. L Devroye, *Non-Uniform Random Variate Generation*. (Springer-Verlag, New York), (1986).
17. S Athey, G Imbens, Recursive partitioning for heterogeneous causal effects. *Proceedings of the National Academy of Sciences* **113**, 7353–7360 (2016).
18. R Durrett, *Probability: Theory and Examples*. (Cambridge University Press), 5th edition, (2019) Page references are to the author's version of January 11, 2019. https://sites.math.duke.edu/~rtd/PTE/PTE5_011119.pdf.
19. R Latała, D Matlak, Royen's proof of the Gaussian correlation inequality in *Geometric Aspects of Functional Analysis: Israel Seminar (GAFA) 2014–2016*, Lecture Notes in Mathematics, eds. B Klartag, E Milman. (Springer International Publishing) Vol. 2169, pp. 265–275 (2017).
20. R Vershynin, *High-Dimensional Probability: An Introduction with Applications in Data Science*. (Cambridge University Press), 1st edition, (2018).
21. J Pitman, M Yor, A guide to Brownian motion and related stochastic processes (2018) arXiv:1802.09679.
22. JK Hunter, Notes on partial differential equations (2014) Lecture notes, University of California, Davis. Revised June 18, 2014. https://www.math.ucdavis.edu/~hunter/pdes/pde_notes.pdf.
23. I Waudby-Smith, M Larsson, A Ramdas, Nonasymptotic and distribution-uniform Komlós–Major–Tusnády approximation (2025) arXiv:2502.06188.
24. DA Levin, Y Peres, *Markov Chains and Mixing Times*. (American Mathematical Society, Providence, RI), 2nd edition, (2017) With contributions by Elizabeth L. Wilmer.
25. W. E. Upjohn Institute for Employment Research, The National JTPA Study (1993) Public-use data documentation. Accessed October 22, 2026. https://www.upjohn.org/data-tools/employment-research-data-center/national-jtpa-study.
26. HS Bloom, et al., The benefits and costs of JTPA Title II-A programs: Key findings from the National Job Training Partnership Act Study. *Journal of Human Resources* **32**, 549–576 (1997).



(a) NSS-IPW **(b)** NSS-DIM **(c)** NSS-SSE

(d) HON-IPW **(e)** HON-DIM **(f)** HON-SSE

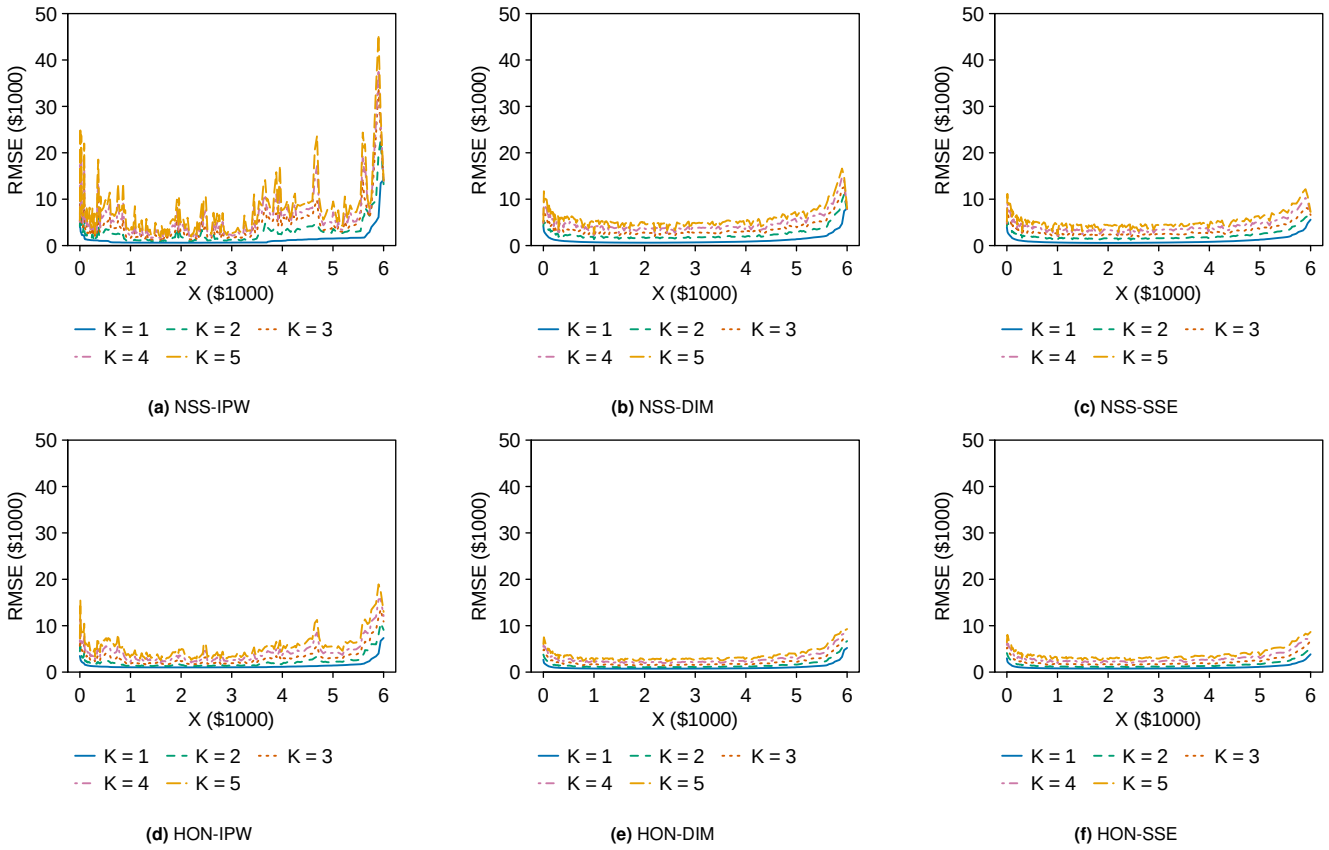
Fig. S1. Monte Carlo evidence for the univariate design ($p = 1$). Panels report pointwise root mean squared error (RMSE) over $[0, 1]$ for tree depths $K = 1, \ldots, 5$, comparing estimators without sample splitting (NSS, top row) and with honest sample splitting (HON, bottom row) across the inverse probability weighting (IPW), difference in means (DIM), and sum of squared errors (SSE) splitting criteria (columns). The covariate X is uniform on $[0, 1]$. Results are based on $1{,}000{,}000$ replications.



(a) NSS-IPW **(b)** NSS-DIM **(c)** NSS-SSE

(d) HON-IPW **(e)** HON-DIM **(f)** HON-SSE

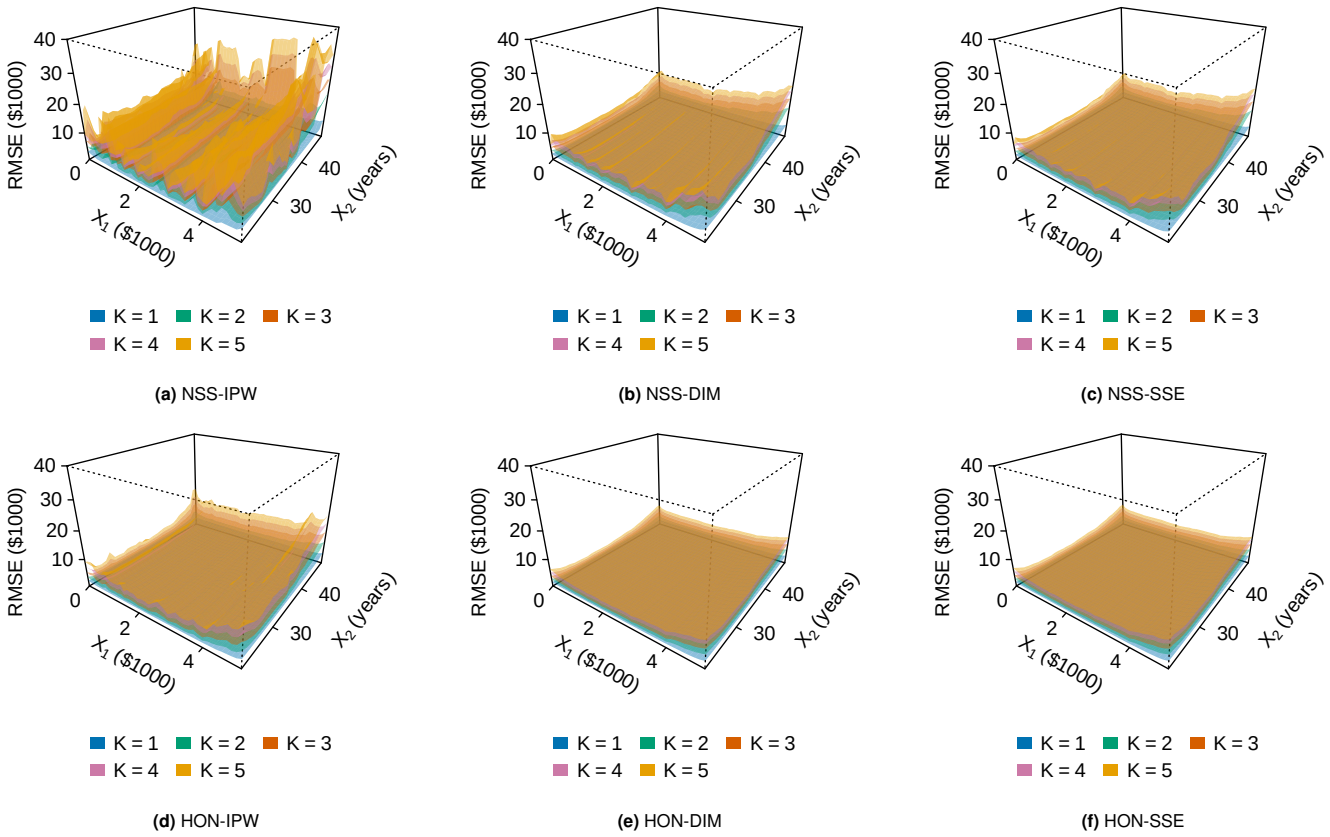
Fig. S2. Monte Carlo evidence for the bivariate design ($p = 2$). Panels report pointwise root mean squared error (RMSE) over $[0, 1]^2$ for tree depths $K = 1, \ldots, 5$, comparing estimators without sample splitting (NSS, top row) and with honest sample splitting (HON, bottom row) across the inverse probability weighting (IPW), difference in means (DIM), and sum of squared errors (SSE) splitting criteria (columns). The covariates X_1 and X_2 are independent and uniform on $[0, 1]$. Results are based on $1{,}000{,}000$ replications.



(a) NSS-IPW **(b)** NSS-DIM **(c)** NSS-SSE

(d) HON-IPW **(e)** HON-DIM **(f)** HON-SSE

Fig. S3. Empirical JTPA resampling evidence for the one-covariate design ($p = 1$). Panels report pointwise RMSE over the empirical support of preprogram earnings (X) for tree depths $K = 1, \ldots, 5$, comparing NSS (top row) and HON (bottom row) across the IPW, DIM, and SSE criteria (columns). Earnings and RMSE are displayed in thousands of dollars. Results are based on $1{,}000{,}000$ replications.



(a) NSS-IPW **(b)** NSS-DIM **(c)** NSS-SSE

(d) HON-IPW **(e)** HON-DIM **(f)** HON-SSE

Fig. S4. Empirical JTPA resampling evidence for the two-covariate design ($p = 2$). Panels report pointwise RMSE on a rectangular evaluation grid spanning the marginal 1st–99th percentiles of preprogram earnings (X_1) and age (X_2) for tree depths $K = 1, \ldots, 5$, comparing NSS (top row) and HON (bottom row) across the IPW, DIM, and SSE criteria (columns). Earnings and RMSE are displayed in thousands of dollars and age in years. Results are based on $1{,}000{,}000$ replications.