Exact Coverage Frontiers for Multi-Environment Jackknife+

Abstract

We answer the question raised by Duchi, Gupta, Jiang, and Sur [5] of replacing the conservative multi-environment jackknife-minmax envelope by corrected jackknife+ endpoint quantiles. Assuming exchangeability across environments and allowing arbitrary within-environment dependence, we characterize exactly when an interval covers at least a $1-\alpha$ fraction of a new environment with probability at least $1-\delta$. The solution jointly calibrates the within-environment residual rank and the across-environment endpoint ranks. For equal-size environments, we obtain the exact worst-case failure probability and all valid finite-sample rank pairs, including rounding and ties. The limiting validity budget is $a/\alpha + 2b/\delta \leq 1$, where a and b are the inner and outer trimming levels. Comparison-graph degree bounds prove validity, and explicit tie-free constructions establish sharpness. For a fixed multiset of unequal environment sizes, an exact Hall-type characterization is computable in quadratic time after sorting when the induced inner ranks and outer rank are positive. In the balanced positive-rank model, asymmetric endpoint ranks offer no minimax improvement, while hidden independent global outer-rank randomization admits an exact tournament-score optimization. We also derive an exact coverage guarantee for an expanded rule under simultaneous on-sample deletion stability and sharp degradation under approximate block exchangeability. Finally, an atomless construction exhibits an arbitrarily large pathwise average-length improvement over jackknife-minmax, demonstrating the potential efficiency gain of calibrated endpoint quantiles.

Note. This paper was generated entirely by AI, including MiMo, using an automated research pipeline developed by Chenghua Liu and Hanyu Li.

1 Introduction and problem statement

Multi-environment predictive inference seeks a block-level guarantee rather than marginal coverage for one response. Given m training environments, the goal is to construct an interval-valued predictor $\mathcal{I}$ such that the realized fraction of responses satisfying $Y \in \mathcal{I}(X)$ in a new environment is at least $1-\alpha$ with probability at least $1-\delta$. Following Duchi et al. [5], we treat each complete environment—including its possibly random positive size—as one exchangeable unit and impose no independence or exchangeability condition within an environment; this formulation isolates what block exchangeability alone can guarantee.

The multi-environment jackknife-minmax method of Duchi et al. [5] obtains this distribution-free guarantee by taking extreme prediction envelopes. Their remark following Theorem 1 raises the possibility of replacing those extremes by suitably corrected lower and upper quantiles, as in jackknife+. In Appendix D (Algorithm 9 and Figure D.11), they report a shorter uncorrected rule with empirical undercoverage. The remark suggests that the arbitrariness of the block model may preclude a correction altogether. We show instead that non-extreme endpoint rules are possible, but only inside an exact two-level rank budget.

The closest benchmarks separate along the exchangeable unit and prediction target. Ordinary jackknife+ supplies the comparison-tournament factor for one response [1]. Batch predictive inference controls a covered fraction under individual-level exchangeability and uncoupled ranks [14], whereas hierarchical jackknife+ targets one new response and uses within-group assumptions [13]. Here whole environments are exchangeable, their coordinates may be arbitrarily dependent, and the endpoint candidates share coupled leave-two-environment fits.

The first level is the familiar jackknife+ comparison tournament. The second is specific to the block-fraction target: different missed coordinates of the test block can be defeated by different subsets of calibration environments. A coordinate–environment incidence count converts these misses into an outdegree demand, and the set of simultaneously satisfiable demands determines the sharp coverage frontier. This viewpoint yields the following results.

- **A closed-form balanced frontier.** For a fixed common block size n, Theorem 3.1 gives the exact finite-sample worst-case failure probability, including all rounding and tie effects. With $N = m + 1$, $F = \lfloor \alpha n \rfloor + 1$, and $q = \lceil \lfloor \delta N \rfloor /2 \rceil$, Corollary 3.2 gives every valid positive finite-rank cell, while Corollary 3.3 shows that the Pareto antichain has exactly $\min\{F, q\}$ points. Its limiting boundary is $a/\alpha + 2b/\delta = 1$.

- **Heterogeneous blocks and explicit extremizers.** The comparison-graph method extends beyond equal-size blocks. Theorem 6.1 gives an exact formula for every fixed multiset of possibly unequal positive block sizes in the finite-rank regime. The formula is a degree-demand feasibility problem computable with $O(N^2)$ unit-cost arithmetic operations after sorting. Matching lower bounds are realized by one additive block-permutation-invariant multiset learner and remain sharp without ties.

- **Asymmetric and randomized ranks.** In the balanced positive-rank model, Theorem 6.2 proves that separate lower and upper outer ranks enter the minimax law only through their maximum. Thus symmetric outer trimming is a conclusion, not an assumption. With a fixed inner rank, Theorem 6.3 exactly characterizes a hidden independent global outer rank by optimization over tournament score sequences; this value can be strictly below the corresponding convex combination, as Proposition 6.4 demonstrates.

- **Robustness, impossibility, and efficiency.** Under simultaneous on-sample deletion stability, the 2ε-expanded rule has the exact scalar outer-rank frontier. The exact pointwise count combines with the swap-TV principle of Barber et al. [2] to give a sharp additive guarantee in the balanced finite-rank setting under approximate block exchangeability. At the negative extreme, the ordinary outer factor-two choice can have failure probability one. At the positive extreme, a valid non-extreme rule has endpoint-to-minmax pathwise average-length ratio tending to zero on an explicit atomless family.

Proof architecture. Reveal all $N = m + 1$ blocks and treat each label in turn as the hypothetical test block. Pairwise leave-two-block residual quantiles define an oriented comparison graph. If one hypothetical test block fails on F coordinates, an incidence count forces that vertex to exceed a rank-dependent outdegree threshold w. A sharp count then bounds the number of vertices with such high outdegree by $\min\{N, 2(N - w) - 1\}$. Exchangeability turns this pointwise label count into a failure probability. For sharpness, an extremal oriented graph is realized by equal-row-sum binary incidence matrices; Gale–Ryser supplies the matrices, and a finite prediction table realizes all directed comparisons simultaneously. An additive type-count representation then turns this table into a symmetric multiset learner. For unequal sizes the vertices have different outdegree demands, whose simultaneous feasibility is exactly Hall's condition. Random outer ranks replace a single degree threshold by a monotone payoff on tournament scores. The graph-count argument is needed only in the nontrivial $1 \le t \le F$ branch; the other upper-bound branches are automatic, and the matching construction directly realizes the $t > F$ case.

2 Setup and deterministic rank conventions

This section fixes the balanced statistical experiment and the admissible learners. It then defines the order-statistic conventions, endpoint rule, block-failure event, and integer ranks used throughout the paper.

2.1 Balanced exchangeable blocks and invariant learners

Let covariates take values in a standard Borel space $\mathcal{X}$, let responses be real-valued, let $m, n \ge 1$, and put $N := m + 1$. Write

$$Z_i = \big((X_j^i, Y_j^i)\big)_{j=1}^n, \qquad i = 1, \dots, N.$$

Let $\mathcal{Z} := (\mathcal{X} \times \mathbb{R})^n$ be the common block space. The random blocks $(Z_1, \dots, Z_N)$ are *exchangeable*: their joint law is invariant under every permutation of the N block labels. No independence or exchangeability is imposed on the n observations within a block. This common-size model is the balanced specialization of Duchi et al. [5]; their general block model permits unequal and random sizes. The exact heterogeneous extension is given in Section 6.1.

For every $r \ge 0$, a learning rule supplies a jointly measurable map

$$\mathcal{A}_r : \mathcal{Z}^r \times \mathcal{X} \longrightarrow \mathbb{R}$$

that is symmetric in its r block arguments. We write $\mathcal{A}(T)(x)$ when the arity is determined by the finite training sequence T. Equivalently, the rule acts on finite multisets: multiplicities are retained, deletion removes one indexed occurrence, and every sum over training blocks counts multiplicity. The main statements concern deterministic rules. Their worst-case values are unchanged for a coherently randomized family $(u, T, x) \mapsto \mathcal{A}_u(T)(x)$, jointly measurable and symmetric for almost every u, when its seed U is independent of $(Z_1, \dots, Z_N)$ and the same realized seed is used in every augmented fit. Conditioning on U gives the deterministic upper bound, while the matching lower bound is deterministic. Independently rerandomized evaluations of the same training multiset are not covered by this coupling.

2.2 Extended order statistics

Responses and learner predictions are assumed to be finite real numbers; extended values enter only through the following convention. Write $\overline{\mathbb{R}} := \mathbb{R} \cup \{-\infty, +\infty\}$. For $u_1, \dots, u_r \in \overline{\mathbb{R}}$, let

$$u_{(1)} \le \cdots \le u_{(r)}$$

denote the order statistics, retaining ties with multiplicity, and set

$$u_{(0)} := -\infty, \qquad u_{(r+1)} := +\infty.$$

For $\eta \in [0, 1)$, define the deterministic lower and upper conformal quantiles

$$q^-_{r,\eta}(u) := u_{(\lfloor \eta(r+1) \rfloor)}, \qquad q^+_{r,\eta}(u) := u_{(r+1-\lfloor \eta(r+1) \rfloor)}. \tag{2.1}$$

We use the extended-real conventions $c-(+\infty)=-\infty$, $c+(+\infty)=+\infty$, $(-\infty)-r=-\infty$, and $(+\infty)+r=+\infty$ for finite $c\in\mathbb{R}$ and $r\geq 0$. Throughout, $[L,U]:=\{y\in\mathbb{R}: L\leq y\leq U\}$; hence it is empty when $L>U$, and equality with an endpoint counts as coverage. The notation $[L,U]^{+r}:=[L-r,U+r]$ means endpoint inflation, interpreted by the same convention even when the original endpoints cross. Let $\mathrm{len}(I)$ be one-dimensional Lebesgue measure, with $\mathrm{len}(\varnothing)=0$ and $\mathrm{len}(I)=+\infty$ for a nonempty unbounded interval. For an interval-valued rule I, its pathwise average length on a realized test block is

$$\frac{1}{n}\sum_{j=1}^{n}\mathrm{len}\big(I(X_j^N)\big).$$

2.3 The endpoint-quantile rule

Given training blocks $Z_1,\dots,Z_m$, let

$$\widehat{f}_{-i}:=\mathcal{A}(Z_1,\dots,Z_{i-1},Z_{i+1},\dots,Z_m)$$

and define the within-environment score radius

$$S_i(a):=q^{+}_{n,a}\left(\left(\left|Y_j^i-\widehat{f}_{-i}(X_j^i)\right|\right)_{j=1}^{n}\right). \tag{2.2}$$

For a covariate x, the two-level endpoint rule is

$$\mathcal{I}_{a,b}(x):=\left[q^{-}_{m,b}\big((\widehat{f}_{-i}(x)-S_i(a))_{i=1}^{m}\big),q^{+}_{m,b}\big((\widehat{f}_{-i}(x)+S_i(a))_{i=1}^{m}\big)\right]. \tag{2.3}$$

If the displayed lower endpoint exceeds the upper endpoint, the interval is understood to be empty. In the only regime where a nontrivial upper bound is needed, Lemma 4.1 proves that this cannot occur.

For $\alpha\in(0,1)$, the new block fails when

$$\frac{1}{n}\sum_{j=1}^{n}\mathbf{1}\{Y_j^N\in\mathcal{I}_{a,b}(X_j^N)\}<1-\alpha. \tag{2.4}$$

To place the minimax value on one fixed experiment, take the universal covariate space $\mathcal{X}_\star:=[0,1]$ and let $\mathcal{Z}_\star:=(\mathcal{X}_\star\times\mathbb{R})^n$. The exact worst-case failure probability is

$$\Gamma_{m,n,\alpha}(a,b):=\sup_{\mathsf{P},\mathcal{A}}\mathbb{P}_{\mathsf{P}}\{(2.4)\text{ holds}\}, \tag{2.5}$$

where P ranges over exchangeable laws on $\mathcal{Z}_\star^N$ and $\mathcal{A}$ over the measurable symmetric learners defined above. Every upper bound below holds for any fixed standard Borel covariate space $\mathcal{X}$. The lower bound needs only the finite space $\{0,\dots,N-1\}\times\{1,\dots,n\}$, and therefore embeds into any standard Borel covariate space containing at least Nn distinct points.

Problem 2.1 (Two-level rank correction addressed here)**.** For fixed m,n,α, determine $\Gamma_{m,n,\alpha}(a,b)$ exactly, including finite-sample rounding and ties. For every $\delta\in(0,1)$, characterize all deterministic rank pairs for which $\Gamma_{m,n,\alpha}(a,b)\leq\delta$, and decide whether some valid non-extreme pair has strictly smaller pathwise average interval length across the n test coordinates than jackknife-minmax on a nondegenerate exchangeable class.

For a fixed $\delta\in(0,1)$, the minmax interval from Algorithm 1 of Duchi et al. [5], used later for comparison, can be written at x as

$$\mathcal{I}_{\mathrm{mm}}(x)=\left[\min_{1\leq i\leq m}\widehat{f}_{-i}(x)-Q_\delta,\ \max_{1\leq i\leq m}\widehat{f}_{-i}(x)+Q_\delta\right],\quad Q_\delta:=q^{+}_{m,\delta}\big((S_i(\alpha))_{i=1}^{m}\big). \tag{2.6}$$

2.4 Integer ranks

Fix $a, b \in [0,1)$ and set

$$F := \lfloor \alpha n \rfloor + 1, \qquad t := \lfloor a(n+1) \rfloor, \qquad \ell := \lfloor bN \rfloor. \tag{2.7}$$

Thus block failure means at least F misses; the inner radius in (2.2) has ascending rank $n+1-t$; and the outer lower and upper endpoints in (2.3) have ranks ℓ and $N-\ell$, respectively, among $m = N-1$ candidates.

If $t = 0$, every inner radius is $+\infty$, while if $\ell = 0$, the outer endpoints are $-\infty$ and $+\infty$. Under (2.1), either case yields

$$\Gamma_{m,n,\alpha}(a,b) = 0. \tag{2.8}$$

Without extended order statistics, these rank-zero choices should instead be called undefined; they are not finite-rank rules. The remainder of the main theorem concerns

$$1 \le t \le n, \qquad 1 \le \ell \le m. \tag{2.9}$$

3 Exact failure law and rank frontier

The next theorem identifies the exact minimax cost of combining the inner and outer ranks, thereby resolving problem 2.1. Its two branches separate an overaggressive inner rank from the incidence-count regime.

Theorem 3.1 (Exact worst-case failure law)**.** *Assume* (2.9)*. If $1 \le t \le F$, define*

$$L := \min\left\{m, \left\lfloor \frac{F(\ell-1)}{F-t+1} \right\rfloor\right\}. \tag{3.1}$$

Then

$$\Gamma_{m,n,\alpha}(a,b) = \begin{cases} 1, & t > F, \\ \min\left\{1, \dfrac{2L+1}{N}\right\}, & 1 \le t \le F. \end{cases} \tag{3.2}$$

The upper bound holds uniformly over every admissible exchangeable law and learner. For every parameter tuple, a single measurable block-permutation-invariant learner and a finite exchangeable law attain equality. The attaining construction may moreover be perturbed so that all residual, comparison, endpoint, and response-on-endpoint ties are absent.

For the proof, put

$$d := N - \ell, \qquad D := F - t + 1, \tag{3.3}$$

and define

$$w := \begin{cases} 0, & t > F, \\ \max\left\{0, \left\lceil \dfrac{Fd - m(t-1)}{F-t+1} \right\rceil\right\}, & 1 \le t \le F. \end{cases} \tag{3.4}$$

The theorem is equivalently

$$\Gamma_{m,n,\alpha}(a,b) = \frac{\min\{N,\, 2(N-w)-1\}}{N}. \tag{3.5}$$

Indeed, when $t \le F$,

$$Fd - m(t-1) = m(F-t+1) - F(\ell-1) = mD - F(\ell-1), \tag{3.6}$$

so the elementary identity $\lceil m - x \rceil = m - \lfloor x \rfloor$, together with the truncations in (3.1) and (3.4), gives

$$w = m - L. \tag{3.7}$$

3.1 Exact lattice Pareto frontier

We first translate the exact law into a criterion at a target block-failure level δ.

Corollary 3.2 (Finite-sample validity criterion)**.** *Let $\delta \in (0,1)$, and define*

$$D_\delta := \lfloor \delta N \rfloor, \qquad q := \left\lceil \frac{D_\delta}{2} \right\rceil. \tag{3.8}$$

If $D_\delta = 0$, no finite-rank endpoint rule satisfies $\Gamma_{m,n,\alpha}(a,b) \le \delta$. If $D_\delta \ge 1$, a finite rank pair is valid if and only if

$$1 \le t \le F, \qquad 1 \le \ell \le q, \qquad F(q-\ell+1) \ge q(t-1)+1. \tag{3.9}$$

Proof. Every finite-rank rule has failure probability at least $1/N$, proving the first assertion. Now $D_\delta \le N-1$, so Theorem 3.1 gives

$$\Gamma_{m,n,\alpha}(a,b) \le \delta \iff 2L+1 \le D_\delta \iff L \le q-1.$$

Because $q-1 < m$, the last inequality removes the truncation in (3.1) and is equivalent to

$$F(\ell-1) \le q(F-t+1)-1.$$

Rearranging gives the final inequality in (3.9); it also forces $\ell \le q$. The condition $t \le F$ is necessary because the alternative has failure probability one. $\square$

For the remainder of this subsection assume $D_\delta \ge 1$, and hence $q \ge 1$. For each $\ell = 1,\dots,q$, define

$$T_\ell := 1 + \left\lfloor \frac{F(q-\ell+1)-1}{q} \right\rfloor. \tag{3.10}$$

Increasing either t or ℓ can only shrink the endpoint interval. The sequence T_ℓ is nonincreasing, and the maximal valid rank pairs are

$$(T_\ell, \ell), \qquad \ell = 1,\dots,q,$$

after retaining only the last pair on each plateau $T_\ell = T_{\ell+1}$. These remaining pairs form the exact lattice Pareto antichain. The continuous parameters merely label left-closed, right-open rank cells:

$$\frac{t}{n+1} \le a < \frac{t+1}{n+1}, \qquad \frac{\ell}{N} \le b < \frac{\ell+1}{N}. \tag{3.11}$$

For fixed data and covariate x, the interval depends on (a,b) only through (t,ℓ) and is therefore constant within each cell.

The apparently nested floor in (3.10) has a useful closed form.

Corollary 3.3 (Size and enumeration of the Pareto antichain)**.** *For $D_\delta \ge 1$,*

$$T_\ell = \left\lceil \frac{F(q-\ell+1)}{q} \right\rceil. \tag{3.12}$$

After dominated plateaus are removed, the exact lattice Pareto antichain has cardinality

$$\min\{F, q\}.$$

It can therefore be enumerated with $O(\min\{F,q\})$ arithmetic operations. If $F \le q$, an explicit enumeration is

$$\left(t,\ q - \left\lfloor \frac{q(t-1)}{F} \right\rfloor\right), \qquad t = 1,\dots,F. \tag{3.13}$$

If $F \ge q$, it is

$$\left(\left\lceil \frac{Fr}{q} \right\rceil,\ q - r + 1\right), \qquad r = 1,\dots,q. \tag{3.14}$$

Proof. For positive integers x, q, $1 + \lfloor (x-1)/q \rfloor = \lceil x/q \rceil$, which proves (3.12). Put $r = q - \ell + 1$. If $F \ge q$, the sequence $\lceil Fr/q \rceil$, $r = 1,\dots,q$, is strictly increasing and hence has q distinct values. If $F < q$, consecutive values differ by at most one, the first value is one, and the last is F; it therefore takes every value in $\{1,\dots,F\}$. Retaining the largest ℓ on each plateau gives (3.13); there are no plateaus when $F \ge q$, giving (3.14). □

3.2 Asymptotic budget

We next pass from the finite rank lattice to a continuous budget by letting both block counts diverge.

Corollary 3.4 (Asymptotic frontier)**.** *Let $\min\{m,n\} \to \infty$ along any integer sequence and fix $a, b, \alpha, \delta \in (0,1)$. Then*

$$\Gamma_{m,n,\alpha}(a,b) \longrightarrow \begin{cases} \min\left\{1, \dfrac{2\alpha b}{\alpha - a}\right\}, & a < \alpha, \\ 1, & a \ge \alpha. \end{cases} \tag{3.15}$$

Consequently, the limiting failure probability is at most δ exactly on the side

$$\frac{a}{\alpha} + \frac{2b}{\delta} \le 1, \tag{3.16}$$

with Pareto boundary given by equality.

Proof. If $a < \alpha$, integer rounding is asymptotically negligible, while the truncation in (3.1) gives

$$\frac{L}{m} \longrightarrow \min\left\{1, \frac{\alpha b}{\alpha - a}\right\}.$$

Substitution into (3.2) gives the first case. If $a > \alpha$, eventually $t > F$. At $a = \alpha$, one has $t \le F$ and $F - t + 1 \in \{1, 2\}$; moreover $F(\ell - 1)/m \to \infty$, so eventually $L = m$. This proves (3.15), and solving its first branch for a failure probability at most δ gives (3.16). □

Thus the conventional outer allocation $b = \delta/2$ exhausts the entire asymptotic budget: among positive finite-rank cells it forces $a \downarrow 0$, and every fixed $a > 0$ lies on the invalid side. Conversely, if $a = \alpha$, then every fixed $b > 0$ has worst-case failure probability converging to one.

Remark 3.5 (Three boundary checks). When $t = 1$, (3.1) gives $L = \ell - 1$, so the law reduces to the ordinary outer tournament factor $\min\{1, (2\ell - 1)/N\}$; in particular this covers $n = 1$. When $\ell = 1$ and $t \le F$, the smallest finite failure probability is $1/N$. When $t > F$, the worst-case failure probability is one. The intercepts $a = 0$ and $b = 0$ of the asymptotic line describe the closure of positive finite-rank cells; literally setting either rank to zero invokes the infinite-endpoint convention (2.8).

The exact law also identifies the entire region where endpoint quantiles can be defeated on every possible test label, rather than merely fail a prescribed δ target.

Corollary 3.6 (Complete-failure phase)**.** *In the finite-rank regime, $\Gamma_{m,n,\alpha}(a,b) = 1$ if and only if either $t > F$, or $t \le F$ and*

$$F(\ell - 1) \ge (F - t + 1)\left\lceil \frac{m}{2} \right\rceil. \tag{3.17}$$

For fixed positive a, b, α, the limiting failure probability in (3.15) *equals one exactly when*

$$a \ge \alpha, \qquad or \qquad a < \alpha \quad and \quad \frac{a}{\alpha} + 2b \ge 1. \tag{3.18}$$

Proof. For $t \le F$, (3.2) equals one exactly when $2L + 1 \ge N = m + 1$, or $L \ge \lceil m/2 \rceil$. Substituting (3.1) gives (3.17). The asymptotic claim follows by setting the first branch of (3.15) equal to one. $\square$

4 Upper bound

The upper bound augments the observed sample by treating every block label as the possible test block. An incidence argument turns coordinate misses into outdegree requirements, a graph count limits the number of such labels, and exchangeability converts that pointwise count into probability.

Reveal all N blocks. For distinct labels i, k, let

$$\widehat{f}_{-\{i,k\}} := \mathcal{A}\big((Z_r)_{r \notin \{i,k\}}\big)$$

be the single predictor associated with the unordered omitted pair $\{i, k\}$, and define

$$R_{ik} := q^{+}_{n,a}\left(\left(\left|Y^i_j - \widehat{f}_{-\{i,k\}}(X^i_j)\right|\right)_{j=1}^{n}\right). \tag{4.1}$$

If i is treated as the hypothetical test block, candidate $k \ne i$ has center and endpoints

$$\begin{aligned} p^i_{kj} &:= \widehat{f}_{-\{i,k\}}(X^i_j), \\ A^i_{kj} &:= p^i_{kj} - R_{ki}, \qquad\qquad B^i_{kj} := p^i_{kj} + R_{ki}. \end{aligned} \tag{4.2}$$

Let $A^i_{(r)j}$ and $B^i_{(r)j}$ denote the separate order statistics over $k \ne i$. The hypothetical interval is

$$\mathcal{I}^i_j = [A^i_{(\ell)j}, B^i_{(d)j}], \qquad d = N - \ell. \tag{4.3}$$

Define the hypothetical failure indicator

$$H_i(Z_1, \ldots, Z_N) := \mathbf{1}\left\{\sum_{j=1}^{n} \mathbf{1}\{Y^i_j \notin \mathcal{I}^i_j\} \ge F\right\}. \tag{4.4}$$

For a permutation σ, define the relabeling action by $(\sigma \cdot Z)_r := Z_{\sigma^{-1}(r)}$. Block-permutation invariance of the learner implies the equivariance identity

$$H_{\sigma(i)}(\sigma \cdot Z) = H_i(Z) \tag{4.5}$$

for every i and σ.

Orient a strict comparison edge by

$$i \longrightarrow k \quad \Longleftrightarrow \quad R_{ik} > R_{ki}. \tag{4.6}$$

This is an oriented graph: each unordered pair carries at most one directed edge, and a tie carries none.

The following structural observation rules out crossed endpoints in the only regime where the graph count improves on the bound one.

Lemma 4.1 (Aggregate interval is nonempty in the nontrivial regime)**.** *Assume* $1 \le t \le F$ *and*

$$w > \left\lfloor \frac{N-1}{2} \right\rfloor.$$

Then $w \le d$*, hence* $\ell \le d$*, and every interval in* (4.3) *is nonempty.*

Proof. Since $d \le m$ and $D = F - t + 1 > 0$,

$$Fd - m(t-1) \le Fd - d(t-1) = dD.$$

The definition (3.4) therefore gives $w \le d$. Since $w > (N-1)/2$ and d is an integer, $2d \ge N$, or equivalently $\ell = N - d \le d$.

For each candidate, $A^i_{kj} \le B^i_{kj}$. Coordinatewise monotonicity of order statistics gives

$$A^i_{(\ell)j} \le B^i_{(\ell)j} \le B^i_{(d)j}.$$

Thus the aggregate lower endpoint does not exceed the aggregate upper endpoint. □

The next lemma is the core link from F missed coordinates to pairwise residual comparisons. Unlike the preceding observation, it also applies when the aggregate endpoints cross.

Lemma 4.2 (Failure forces high outdegree)**.** *Assume the positive finite-rank conditions* (2.9)*, and let* w *be defined by* (3.4)*. If* $H_i = 1$*, then the outdegree* z_i *of* i *in* (4.6) *satisfies*

$$z_i \ge w.$$

Proof. If $t > F$, then $w = 0$, so the conclusion follows from $z_i \ge 0$. Suppose henceforth that $1 \le t \le F$. Choose a set J_i of exactly F missed coordinates. At any $j \in J_i$, the interval convention above implies either

$$Y^i_j < A^i_{(\ell)j} \quad \text{or} \quad Y^i_j > B^i_{(d)j}.$$

In the first case, at least $m - \ell + 1 = d$ candidate lower endpoints exceed the response. In the second, at least d candidate upper endpoints lie below it. Hence at least d of the m candidate intervals miss at each selected coordinate.

For $k \ne i$, let

$$e_k := \sum_{j \in J_i} \mathbf{1}\{Y^i_j \notin [A^i_{kj}, B^i_{kj}]\}.$$

Then $\sum_{k \ne i} e_k \ge Fd$. A candidate miss means

$$\left|Y^i_j - p^i_{kj}\right| > R_{ki}.$$

If $e_k \ge t$, at least t of the n residuals defining R_{ik} are strictly larger than R_{ki}. Since R_{ik} has ascending rank $n + 1 - t$,

$$e_k \ge t \quad \Longrightarrow \quad R_{ik} > R_{ki}, \tag{4.7}$$

so $i \to k$.

The z_i outgoing columns contain at most F incidences each, while every other column contains at most $t - 1$. Therefore

$$Fd \le z_i F + (m - z_i)(t-1) = m(t-1) + z_i(F - t + 1).$$

Taking the integer ceiling and applying (3.4) yields $z_i \ge w$. □

It remains to bound how many vertices can meet this degree demand.

Lemma 4.3 (Sharp oriented-graph count)**.** *Let $N \geq 2$ and $w \in \{0, \ldots, N-1\}$. In an oriented graph on N vertices, at most*

$$s := \min\{N, 2(N-w) - 1\}$$

vertices can have outdegree at least w.

Proof. Let G be the set of such vertices and put $g := |G|$. The case $g = 0$ is immediate. The total outdegree from G is at least gw. There are at most $g(N-g)$ directed edges from G to its complement, and an oriented graph has at most $g(g-1)/2$ edges inside G. Hence

$$gw \leq g(N-g) + \frac{g(g-1)}{2}.$$

Dividing by g and rearranging gives $g \leq 2(N-w) - 1$; also $g \leq N$. □

Proof of the upper bound in Theorem 3.1. If $t > F$, then $w = 0$ and $s = N$, so the pointwise bound follows from $H_i \leq 1$. Suppose $t \leq F$. If $w \leq \lfloor (N-1)/2 \rfloor$, then again $s = N$, and the same argument applies. Otherwise, Lemma 4.2 shows that every failing label has outdegree at least w, and Lemma 4.3 gives the pointwise inequality

$$\sum_{i=1}^{N} H_i \leq s. \tag{4.8}$$

When $i = N$, the leave-two fit $\widehat{f}_{-\{N,k\}}$ is exactly the leave-k-out fit based on the observed blocks and $R_{kN} = S_k(a)$. Consequently, H_N is precisely the indicator of the original failure event (2.4). By exchangeability and (4.5), all H_i have the same expectation. Therefore

$$\mathbb{P}\{H_N = 1\} = \mathbb{E}H_N = \frac{1}{N}\mathbb{E}\left[\sum_{i=1}^{N} H_i\right] \leq \frac{s}{N}.$$

Equations (3.5) and (3.7) give the claimed upper bound. All graph comparisons are strict. Ties can only delete edges, and hence cannot invalidate any step. □

5 Matching construction

We now realize every inequality in the upper bound with one block-permutation-invariant learner. Sharpness requires both an oriented graph attaining the vertex count and coordinate-incidence patterns attaining its degree requirements; the next two lemmas supply these pieces.

Lemma 5.1 (Extremal oriented graph)**.** *Let $N \geq 2$ and $w \in \{0, \ldots, N-1\}$. There is an oriented graph on N vertices containing a set S of*

$$s = \min\{N, 2(N-w) - 1\}$$

vertices, each with outdegree exactly w.

Proof. If $w \le \lfloor (N-1)/2 \rfloor$, then $s = N$. On the cyclic group $\mathbb{Z}/N\mathbb{Z}$, orient

$$u \longrightarrow u + r, \qquad r = 1, \dots, w.$$

There are no reciprocal arcs, and every vertex has outdegree w.

If $w > \lfloor (N-1)/2 \rfloor$, then $s = 2(N-w) - 1$ is odd. Take any set S of size s, orient every edge from S to its complement, and place a regular tournament on S. Each $u \in S$ has outdegree

$$(N-s) + \frac{s-1}{2} = w.$$

Pairs inside the complement may remain unoriented. □

After fixing the graph, we realize at each selected vertex the required number of coordinate misses without creating extra outgoing comparisons.

Lemma 5.2 (Incidence realization)**.** *For each $u \in S$, index the m columns by the vertices $v \neq u$, and mark the w columns indexed by the outneighbors of u. There exists an $F \times m$ binary matrix B^u such that*

1. *every row has sum $d = N - \ell$;*
2. *every marked column has sum at least t;*
3. *every unmarked column has sum at most $t - 1$.*

Proof. If $t > F$, then $w = 0$. Any $F \times m$ binary matrix with all row sums d works, because every column sum is at most $F < t$.

Suppose $t \le F$. Since $Fd - m(t-1) \le d(F - t + 1)$, the definition of w gives $w \le d$, whether or not the final probability bound is nontrivial. Hence

$$wt \le dF = Fd.$$

The definition of w also gives

$$Fd \le wF + (m-w)(t-1).$$

Starting with degree t in every marked column and zero in every unmarked column, add units without exceeding F on marked columns or $t-1$ on unmarked columns. The two inequalities show that integer column degrees $c_1, \dots, c_m$ with total Fd can be chosen in the required ranges.

It remains to realize these degrees with all F row degrees equal to d. Sort $c_1 \ge \cdots \ge c_m$. The Gale–Ryser inequalities hold:

$$\sum_{j=1}^{k} c_j \le \begin{cases} Fk = F\min\{k,d\}, & k \le d, \\ Fd = F\min\{k,d\}, & k > d. \end{cases}$$

Together with $\sum_j c_j = Fd$, the Gale–Ryser theorem [7, 16] yields the desired binary matrix. □

Proof of the lower bound in Theorem 3.1. The construction proceeds in four steps: choose the extremal graph and its incidence matrices, encode their comparisons in a finite prediction table, realize that table with one multiset learner, and average the resulting label failures under a uniform type permutation.

Use the graph from Lemma 5.1 and the matrices from Lemma 5.2. Take the covariate space to be the finite discrete space

$$\mathcal{X}_0 := \{0, \ldots, N-1\} \times \{1, \ldots, n\}.$$

Choose an injection $\mathcal{X}_0 \hookrightarrow \mathcal{X}_\star$ and identify $\mathcal{X}_0$ with its image. Create N deterministic block types

$$(X_j^u, Y_j^u) = ((u, j), 0), \qquad u = 0, \ldots, N-1, \quad j = 1, \ldots, n. \tag{5.1}$$

Uniformly permute these N fixed types among the N block positions. The resulting finite law is block exchangeable.

We first specify the pairwise prediction table. For $u \in S$ and $j \leq F$, set

$$p_{uvj} := \begin{cases} 1, & B_{jv}^u = 0, \\ 3, & B_{jv}^u = 1 \text{ and } v \to u, \\ 2, & B_{jv}^u = 1 \text{ and } v \not\to u. \end{cases} \tag{5.2}$$

Set $p_{uvj} = 1$ when $j > F$ or $u \notin S$. The predictor associated with the missing pair $\{u, v\}$ takes value p_{uvj} at (u, j) and p_{vuj} at (v, j); these are different covariates, so the two assignments create no coupling conflict.

The table can be realized by an additive type-count learner. For a block Z, let $\chi_v(Z)$ indicate that its covariate tuple is $((v, 1), \ldots, (v, n))$; this ignores the responses. Put $c_{uj} := \sum_{v \neq u} p_{uvj}$, and, for a finite training multiset T, define at the designated covariates

$$\mathcal{A}(T)(u, j) := c_{uj} - \sum_{Z \in T} \sum_{v \neq u} \chi_v(Z) p_{uvj}, \tag{5.3}$$

and define it to be zero elsewhere. If T contains precisely the $N-2$ types other than u, v, then (5.3) telescopes to p_{uvj}. The evaluation map is measurable, additive in block-type indicators, and invariant to the ordering of T. The table has $O(N^2 n)$ entries. After decoding a block type from its complete covariate tuple, evaluation is linear in the number of training blocks; no finite permutation orbit is stored.

We first compute its inner radii. For every $u \neq v$,

$$R_{uv} = \begin{cases} 2, & u \to v, \\ 1, & u \not\to v. \end{cases} \tag{5.4}$$

Indeed, if $u \to v$, then $v \not\to u$ because the graph is oriented. Column v of B^u is marked, contains at least t ones, and (5.2) equals 2 on those entries and 1 elsewhere. If $u \not\to v$, that column is unmarked and has at most $t-1$ entries exceeding 1, whether their value is 2 or 3. The $(n+1-t)$th order statistic is therefore respectively 2 or 1. The same conclusion holds when $t > F$, since every column then has fewer than t nonunit entries. If $u \notin S$, which can occur only in the proper-subset graph construction, then u has no outneighbors and $p_{uvj} = 1$ for every v, j; hence $R_{uv} = 1$, again agreeing with (5.4).

Now take $u \in S$ as the hypothetical test type. Candidate v uses radius R_{vu}. For $j \leq F$,

$$p_{uvj} - R_{vu} = \begin{cases} 1, & B_{jv}^u = 1, \\ \leq 0, & B_{jv}^u = 0. \end{cases} \tag{5.5}$$

Each row of B^u has d ones, leaving exactly $m - d = \ell - 1$ nonpositive lower candidates. The ℓth aggregate lower endpoint is therefore 1, while the response is 0. The first F coordinates miss, and

$F/n > \alpha$. At coordinates $j > F$, all predictions equal 1 and all radii are at least 1, so the response is covered.

If $u \notin S$, every prediction at its covariates equals 1, again with radii at least 1, and every response is covered. Thus exactly the s types in S fail. Since the actual test position receives a uniformly random type,

$$\mathbb{P}\{H_N = 1\} = \frac{s}{N},$$

which matches the upper bound. □

5.1 Tie-free sharpness

The discrete extremizer contains many equalities, so we verify that ties are not responsible for sharpness.

Proposition 5.3 (Quantitative removal of ties)**.** *The matching construction can be chosen with no relevant ties. More precisely, fix $0 < \eta < 1/4$. There are perturbations of every finite response and every specified prediction-table entry, each of magnitude at most η, that remove all residual ties, all equalities $R_{uv} = R_{vu}$, all candidate-endpoint ties, and every equality between a response and an aggregate endpoint, while preserving all s failing labels.*

Proof. We first show that sufficiently small perturbations preserve every strict failure margin, and then choose the perturbations in finite general position to remove all relevant equalities.

A response or prediction perturbation of size at most η changes each absolute residual by at most 2η. An order statistic is one-Lipschitz in the sup norm, so each inner radius changes by at most 2η. A candidate endpoint consequently changes by at most 3η, and so does its outer order statistic. In the construction, the failing lower endpoint exceeds the response by one. After perturbation the gap is at least

$$1 - 3\eta - \eta = 1 - 4\eta > 0.$$

Hence every label in S still fails.

Choose the finitely many primitive response and table-entry perturbations and recompute the constants c_{uj} in (5.3). Choose the perturbations so that, together with 1, they are linearly independent over $\mathbb{Q}$; such choices are dense in every sufficiently small cube. Since all unperturbed specified predictions are at least one, $\eta < 1/4$ fixes the sign of every residual. After the residual orderings are selected, every undesired tie listed in the proposition would therefore be an affine relation with integer coefficients among 1 and the primitive perturbations. Each relation is nontrivial: residuals and radii from opposite directions use different ordered-pair/coordinate table entries, distinct candidates have distinct center entries, and a response–endpoint equality contains a center entry absent from the response. Linear independence rules out all such relations simultaneously. The additive definition and the random permutation of fixed block types preserve measurability, block-permutation invariance, and block exchangeability. Finally, the upper bound forbids any additional failing labels once the original s have been retained. □

6 Extensions of the exact frontier

The closed form in Theorem 3.1 is only the most symmetric instance of the comparison-graph method. This section treats unequal block sizes, asymmetric and randomized outer ranks, simultaneous on-sample deletion stability, and approximate exchangeability, in that order.

6.1 Exact frontier for unequal block sizes

For a standard Borel covariate space $\mathcal{X}$, define the variable-length block space

$$\mathcal{Z}_{\mathrm{var}}(\mathcal{X}) := \bigsqcup_{r\geq 1}(\mathcal{X}\times\mathbb{R})^r. \tag{6.1}$$

For an augmented array $Z = (Z_1, \ldots, Z_N)$, write $n_i := |Z_i|$ for the realized size at label i and

$$\mathcal{M}(Z) := \{\{n_1, \ldots, n_N\}\} \tag{6.2}$$

for its unordered size multiset. Fix a multiset $\boldsymbol{n}$ of N positive integers, and let $\mathcal{P}_{\boldsymbol{n}}$ be the exchangeable laws on $\mathcal{Z}_{\mathrm{var}}(\mathcal{X}_\star)^N$ satisfying $\mathbb{P}\{\mathcal{M}(Z) = \boldsymbol{n}\} = 1$. The assignment of sizes to labels may therefore be random. Learners in this subsection are the variable-length analogues of the maps in Section 2: for every $r \geq 0$,

$$\mathcal{A}_r : \mathcal{Z}_{\mathrm{var}}(\mathcal{X}_\star)^r \times \mathcal{X}_\star \longrightarrow \mathbb{R}$$

is jointly measurable and symmetric in its r block arguments.

For the realized label sizes, define

$$F_i := \lfloor \alpha n_i \rfloor + 1, \qquad t_i := \lfloor a(n_i+1) \rfloor, \qquad \ell := \lfloor bN \rfloor, \qquad d := N - \ell. \tag{6.3}$$

Throughout this subsection assume the genuinely finite inner-rank regime

$$1 \leq t_i \leq n_i \quad \text{for every } i, \qquad 1 \leq \ell \leq m. \tag{6.4}$$

This positivity condition matters: if some, but not all, t_i vanish, the infinite radii couple candidate availability across vertices, so the vertexwise degree formula below no longer applies.

The heterogeneous endpoint rule uses the label-specific inner score

$$S_i(a) := q^+_{n_i,a}\left(\left(\left|Y^i_j - \widehat{f}_{-i}(X^i_j)\right|\right)_{j=1}^{n_i}\right) \tag{6.5}$$

in (2.3). On the revealed augmented array, for distinct i, k, set

$$\begin{aligned} R_{ik} &:= q^+_{n_i,a}\left(\left(\left|Y^i_j - \widehat{f}_{-\{i,k\}}(X^i_j)\right|\right)_{j=1}^{n_i}\right), \\ p^i_{kj} &:= \widehat{f}_{-\{i,k\}}(X^i_j), \qquad A^i_{kj} := p^i_{kj} - R_{ki}, \qquad B^i_{kj} := p^i_{kj} + R_{ki}, \end{aligned} \tag{6.6}$$

and define

$$\mathcal{I}^i_j := [A^i_{(\ell)j}, B^i_{(d)j}], \qquad H_i := \mathbf{1}\left\{\sum_{j=1}^{n_i} \mathbf{1}\{Y^i_j \notin \mathcal{I}^i_j\} \geq F_i\right\}. \tag{6.7}$$

For $i = N$, this is exactly the failure event for the variable-length endpoint rule.

For each size type put

$$w_i := \begin{cases} 0, & t_i > F_i, \\ \max\left\{0, \left\lceil \dfrac{F_i d - m(t_i - 1)}{F_i - t_i + 1} \right\rceil\right\}, & t_i \leq F_i. \end{cases} \tag{6.8}$$

Write $w_{(1)} \leq \cdots \leq w_{(N)}$ for these demands in increasing order and define

$$s^*(\boldsymbol{n}) := \max\left\{ s \in \{0, \ldots, N\} : \sum_{j=s-r+1}^{s} w_{(j)} \leq \frac{r(2N-r-1)}{2} \text{ for every } 1 \leq r \leq s \right\}. \tag{6.9}$$

The constraints are vacuous for $s = 0$. The fixed-multiset minimax value is

$$\Gamma_{m,\boldsymbol{n},\alpha}(a, b) := \sup_{\mathsf{P} \in \mathcal{P}_{\boldsymbol{n}},\mathcal{A}} \mathbb{E}_{\mathsf{P}} H_N, \tag{6.10}$$

where the learner supremum is over the admissible variable-length symmetric maps. Thus all objects in the minimax problem are defined on one sample space.

The functional $s^*(\boldsymbol{n})$ is the largest number of demands that an oriented graph can support. The next theorem shows that it is both necessary and attainable.

Theorem 6.1 (Exact heterogeneous-size frontier)**.** *Under* (6.4)*,*

$$\Gamma_{m,\boldsymbol{n},\alpha}(a, b) = \frac{s^*(\boldsymbol{n})}{N}. \tag{6.11}$$

After an $O(N \log N)$ sort, $s^*(\boldsymbol{n})$ is computable with $O(N^2)$ unit-cost arithmetic operations. Equality is attained by a finite exchangeable law and one additive block-permutation-invariant learner, and all relevant ties can be removed.

More generally, if the block-size multiset is random and the ranks t_i are positive almost surely, then every exchangeable block law and admissible learner obey

$$\mathbb{P}\{H_N = 1 \mid \mathcal{M}\} \le \frac{s^*(\mathcal{M})}{N} \quad \textit{almost surely}, \qquad \mathbb{P}\{H_N = 1\} \le \mathbb{E}\left[\frac{s^*(\mathcal{M})}{N}\right]. \tag{6.12}$$

Proof. The upper bound repeats the incidence argument with vertex-specific demands and then applies Hall-type edge capacity. For the lower bound, a matching supplies the oriented graph, Gale–Ryser supplies each coordinate-incidence matrix, and one variable-length type-count learner realizes all comparisons.

Reveal all blocks and let n_i be the realized size of hypothetical test label i. The incidence proof of Lemma 4.2, with (F, t, w) replaced by (F_i, t_i, w_i), shows that a failing label i has outdegree at least w_i. If $t_i > F_i$, this says only $z_i \ge 0 = w_i$; otherwise the same $F_i d$ incidence count applies even when the aggregate endpoints cross.

Let S be the set of failing labels. For every $A \subseteq S$, all outdegree contributed by vertices in A is carried by an edge incident to A. An oriented graph has at most one arc on each unordered pair, so, with $r = |A|$,

$$\sum_{i \in A} w_i \le \binom{r}{2} + r(N - r) = \frac{r(2N - r - 1)}{2}. \tag{6.13}$$

If $g = |S|$, the sum of the r largest demands among the globally smallest g demands is no larger than the sum of the r largest demands inside S. Thus (6.13) implies that g is feasible in (6.9), and $g \le s^*(\boldsymbol{n})$. The pointwise bound $\sum_i H_i \le s^*(\boldsymbol{n})$, followed by equivariance and exchangeability, proves the upper bound in (6.11). For random $\mathcal{M}$, multiply the pointwise inequality by any bounded nonnegative function $\varphi(\mathcal{M})$. Because $\mathcal{M}$ is invariant under relabeling, exchangeability gives

$$\mathbb{E}[\varphi(\mathcal{M}) H_N] = \frac{1}{N} \mathbb{E}\left[\varphi(\mathcal{M}) \sum_i H_i\right] \le \mathbb{E}\left[\varphi(\mathcal{M}) \frac{s^*(\mathcal{M})}{N}\right].$$

The defining property of conditional expectation yields the first inequality in (6.12); taking expectations yields the second.

For sharpness, let S index the s^* smallest demands. Construct a bipartite graph with w_i demand copies of each $i \in S$ on the left and the $\binom{N}{2}$ unordered vertex pairs on the right; a copy of i is adjacent to the pairs incident to i. For a set of r vertices the union of these neighboring pairs has size $\binom{r}{2} + r(N-r)$. The inequalities in (6.9) are therefore exactly Hall's conditions [10], including subsets containing only some copies of a vertex. A matching assigns w_i distinct incident pairs to each $i \in S$. Orient every assigned pair away from the vertex to which it was matched and leave unassigned pairs unoriented. Every $i \in S$ now has outdegree exactly w_i, while vertices outside S have outdegree zero.

For each $i \in S$, repeat Lemma 5.2 with an $F_i \times m$ matrix, row sum d, and column threshold t_i. The same two capacity inequalities hold because

$$w_i t_i \le F_i d \le w_i F_i + (m - w_i)(t_i - 1). \tag{6.14}$$

Gale–Ryser supplies the matrix. Choose an injection into $\mathcal{X}_\star$ from the finite space

$$\mathcal{X}_0 := \bigsqcup_{i=1}^{N} \{(i,j) : 1 \le j \le n_i\},$$

identify $\mathcal{X}_0$ with its image, and give block type i the complete covariate tuple $((i,1),\dots,(i,n_i))$ and zero responses. For ordered $i \ne v$ and $1 \le j \le F_i$, define p_{ivj} by the three cases in (5.2), using B^i; set $p_{ivj} = 1$ when $F_i < j \le n_i$ or $i \notin S$. Let $\chi_v(Z)$ recognize the full type-v covariate tuple and ignore responses, put $c_{ij} := \sum_{v \ne i} p_{ivj}$, and define

$$\mathcal{A}(T)(i,j) := c_{ij} - \sum_{Z \in T} \sum_{v \ne i} \chi_v(Z) p_{ivj} \quad ((i,j) \in \mathcal{X}_0),$$

with value zero outside $\mathcal{X}_0$. This is one measurable symmetric learner on the variable-length block space, and omitting types i, v makes its value at (i,j) equal p_{ivj}. Exactly the labels in S miss their first F_i coordinates. Uniformly permuting the fixed block types gives failure probability s^*/N. The perturbation proof of Proposition 5.3 is finite and typewise, so it also removes all relevant ties here.

After sorting, prefix sums evaluate all inequalities in (6.9) with $O(N^2)$ unit-cost arithmetic operations. Feasibility is monotone in s, which completes the computational claim. $\square$

When all $n_i = n$, every demand equals the w in (3.4). For a candidate set of size s, the tightest condition in (6.9) is the one with $r = s$, giving $s \le 2(N-w) - 1$. Thus (6.11) reduces exactly to (3.5); the heterogeneous theorem is a strict extension rather than a different bound.

6.2 Asymmetric endpoints: symmetry is minimax optimal

The rule (2.3) uses the same outer rank on both sides. To test whether that symmetry conceals a better tradeoff, let $1 \le \ell_-, \ell_+ \le m$ and define the rank-indexed asymmetric rule

$$\mathcal{I}_{a;\ell_-,\ell_+}(x) := \left[\left(\widehat{f}_{-i}(x) - S_i(a)\right)_{(\ell_-)}, \left(\widehat{f}_{-i}(x) + S_i(a)\right)_{(N-\ell_+)} \right]. \tag{6.15}$$

The order statistics are over $i = 1, \dots, m$, as before. Write $\bar{\ell} := \max\{\ell_-, \ell_+\}$.

Theorem 6.2 (Exact asymmetric-rank law). *In the balanced model and finite inner-rank regime, the worst-case failure probability of* (6.15) *is one when $t > F$. If $1 \le t \le F$, it is*

$$\min\left\{1, \frac{2\bar{L}+1}{N}\right\}, \qquad \bar{L} := \min\left\{m, \left\lfloor \frac{F(\bar{\ell}-1)}{F-t+1} \right\rfloor\right\}. \tag{6.16}$$

Consequently, for every data set and covariate x,

$$\mathcal{I}_{a;\bar\ell,\bar\ell}(x) \subseteq \mathcal{I}_{a;\ell_-,\ell_+}(x), \tag{6.17}$$

while the two rules have the same minimax failure probability. Hence every outer-rank Pareto optimum is symmetric.

Proof. We obtain the upper bound by incidence counting, match it by construction and reflection, and finish with endpoint monotonicity.

Put $\bar d := N - \bar\ell$. At a lower-side miss, at least $N - \ell_- \geq \bar d$ candidate intervals miss the response; at an upper-side miss, at least $N - \ell_+ \geq \bar d$ candidates miss it. The incidence proof therefore applies with $d = \bar d$ and gives exactly the upper bound in (6.16). In the only nontrivial case, the resulting degree threshold exceeds $\lfloor (N-1)/2 \rfloor$. As in Lemma 4.1, it is at most $\bar d$, so $\ell_- \leq \bar\ell \leq \bar d \leq N - \ell_+$; this also proves that the asymmetric aggregate interval is nonempty.

If $\bar\ell = \ell_-$, the matching construction of Section 5, with row sum $\bar d$, produces the required number of lower-side failures. If $\bar\ell = \ell_+$, negate every specified prediction in that construction. Absolute residuals and comparison orientations are unchanged. On a failing coordinate the $\bar d$ selected upper candidates now equal -1, all others are nonnegative, and the response zero lies strictly above the $(N - \ell_+)$th upper candidate. This gives the matching upper-side construction. Every reflected prediction has absolute value at least one, so the fixed-sign, tie-free perturbation argument is unchanged by reflection.

Finally, increasing ℓ_- raises the lower endpoint and increasing ℓ_+ lowers the upper endpoint. Raising the smaller rank to $\bar\ell$ therefore gives (6.17), while (6.16) is unchanged. □

The theorem also clarifies the zero-rank boundary. If both outer ranks vanish, the interval is $\mathbb{R}$; if exactly one vanishes, the rule is one-sided and is not covered by the rank-zero identity (2.8). We restrict Theorem 6.2 to positive ranks precisely to keep these cases separate.

6.3 An exact frontier for a randomized outer rank

Randomly interpolating between integer ranks is not described by the convex hull of their separate worst-case values: the adversary must use one data law and one learner against all ranks in the mixture. We can solve this issue exactly in the balanced model when the inner rank is fixed. For a positive outer rank $\ell \in \{1, \dots, m\}$, define

$$\mathcal{I}_j^{i,(\ell)} := [A^i_{(\ell)j}, B^i_{(N-\ell)j}], \qquad H_i^{(\ell)} := \mathbf{1}\left\{ \sum_{j=1}^{n} \mathbf{1}\{Y_j^i \notin \mathcal{I}_j^{i,(\ell)}\} \geq F \right\}. \tag{6.18}$$

These objects obey the same relabeling equivariance as H_i.

Let π be a distribution on $\{1, \dots, m\}$. The procedure draws one global rank $\mathsf{L} \sim \pi$, independently of the blocks and any learner seed, and uses it for both endpoints, all coordinates, and every hypothetical test label. The law and learner may depend on the known distribution π, but the learner does not observe the realized L; a randomized learner uses the same independent seed in every fit and rank. The corresponding minimax value is

$$\Gamma^{\pi}_{m,n,\alpha}(a) := \sup_{\mathsf{P},\mathcal{A}} \sum_{\ell=1}^{m} \pi_\ell \, \mathbb{E}_{\mathsf{P}} H_N^{(\ell)}, \tag{6.19}$$

with the supremum outside the rank expectation and over the balanced experiment in (2.5). If $t = 0$, every supported rank covers. If $t > F$, a single adversary makes every supported rank fail: use the constant-zero learner and give each block exactly F nonzero responses and $n-F$ zeros. Every inner radius is zero, every positive outer rank returns $[0,0]$, and every label fails. We therefore focus below on $1 \le t \le F$.

Fix $1 \le t \le F$, let $D := F - t + 1$, and for each positive outer rank define

$$w_\ell := \max\left\{0, \left\lceil \frac{F(N-\ell) - m(t-1)}{D} \right\rceil\right\}, \qquad \Phi_\pi(z) := \sum_\ell \pi_\ell \mathbf{1}\{z \ge w_\ell\}. \tag{6.20}$$

Let $\mathsf{Score}(N)$ be the set of tournament outdegree sequences. Equivalently, after sorting $0 \le z_1 \le \cdots \le z_N \le N-1$, Landau's theorem [12] characterizes membership by

$$\sum_{i=1}^k z_i \ge \binom{k}{2} \quad (1 \le k < N), \qquad \sum_{i=1}^N z_i = \binom{N}{2}. \tag{6.21}$$

Theorem 6.3 (Exact randomized-rank law)**.** *In the balanced model, fix $1 \le t \le F$ and a distribution π on $\{1, \dots, m\}$. Under the hidden global independent-rank coupling above,*

$$\Gamma^\pi_{m,n,\alpha}(a) = \frac{1}{N} \max_{(z_1,\dots,z_N)\in\mathsf{Score}(N)} \sum_{i=1}^N \Phi_\pi(z_i). \tag{6.22}$$

The maximum can be computed by dynamic programming with $O(N^4)$ unit-cost arithmetic operations. A single finite exchangeable construction jointly realizes $H_i^{(\ell)} = \mathbf{1}\{z_i \ge w_\ell\}$ for every supported rank under a maximizing tournament, thereby attaining the π-weighted objective; all relevant ties can be removed.

Proof. The proof first bounds every rank-specific failure by a monotone payoff of one comparison-graph degree sequence. A maximizing tournament then drives one common lower construction for all ranks, and Landau's inequalities yield the dynamic program.

For any revealed augmented data set, construct the strict-comparison oriented graph (4.6) and let z_i be its outdegrees. Because every supported rank is positive and finite, Lemma 4.2 applied at rank ℓ gives

$$H_i^{(\ell)} \le \mathbf{1}\{z_i \ge w_\ell\}.$$

Complete every missing edge of the oriented graph in an arbitrary direction. This only increases degrees, and Φ_π is nondecreasing. Averaging over the global independent rank and over hypothetical labels, then applying exchangeability, gives the upper bound in (6.22). The completion need not be measurable or equivariant because it is used only in this pointwise finite optimization.

For the converse, take a maximizing tournament and a vertex of outdegree z. Define

$$c(z) := \left\lfloor \frac{m(t-1) + zD}{F} \right\rfloor. \tag{6.23}$$

Since

$$m(t-1) + zD - zF = (m-z)(t-1) \ge 0,$$

we have $z \le c(z) \le m$, and

$$zt \le Fc(z) \le zF + (m-z)(t-1). \tag{6.24}$$

Thus an $F \times m$ binary incidence matrix exists with every row sum $c(z)$, outneighbor columns of sum at least t, and all other columns of sum at most $t-1$; this is exactly the Gale–Ryser argument of Lemma 5.2.

Use these vertex-dependent matrices in the additive construction of Section 5. For a vertex of degree z_i, each of the first F coordinates has exactly $c(z_i)$ lower candidates equal to one and all others nonpositive, while every upper candidate lies above the response zero. It therefore fails at outer rank ℓ exactly when

$$c(z_i) \geq N - \ell \quad \iff \quad z_i \geq w_\ell. \tag{6.25}$$

Uniformly permuting the block types gives equality in (6.22). This equivalence holds for every $1 \leq \ell \leq m$; noncrossing endpoints are not required. The perturbation of Proposition 5.3 preserves every failure pair (i, ℓ) with positive π_ℓ-mass. Because their weighted total already equals the universal upper bound, no new positive-mass failure can be introduced; hence a tie-free extremizer exists.

For computation, sort a score sequence and let $V_k(r, s)$ be the largest payoff for a nondecreasing length-k prefix with last degree r, degree sum s, and all applicable prefix inequalities in (6.21). Initialize $V_1(r, r) = \Phi_\pi(r)$ for $0 \leq r \leq N - 1$, set all other states to $-\infty$, and for $k \geq 2$ use the recurrence

$$V_k(r, s) = \Phi_\pi(r) + \max_{r' \leq r} V_{k-1}(r', s - r).$$

Prefix maxima make each transition constant time. There are $O(N^4)$ states over all layers and $O(N^3)$ rolling storage; imposing total sum $\binom{N}{2}$ at the final layer proves the arithmetic-operation claim. $\square$

Proposition 6.4 (Randomization is not convexification)**.** *Let $m = 6$, $n = 1$, and $\alpha = a = 1/2$, so $N = 7$ and $F = t = 1$. Let π be uniform on outer ranks two and three. Then*

$$\Gamma^{\pi}_{6,n,\alpha}(a) = \frac{3}{7} < \frac{1}{2}\left(\frac{3}{7} + \frac{5}{7}\right) = \frac{4}{7}, \tag{6.26}$$

where $3/7$ and $5/7$ are the separate deterministic minimax values.

Proof. Here $w_2 = 5$, $w_3 = 4$, and $\Phi_\pi(z) = \mathbf{1}\{z \geq 5\} + \frac{1}{2}\mathbf{1}\{z = 4\}$. Put $A = \#\{i : z_i \geq 5\}$ and $B = \#\{i : z_i \geq 4\}$. The sharp single-threshold count gives $A \leq 3$ and $B \leq 5$. If $A + B \geq 7$, either $A = 3, B \geq 4$, in which case the four largest degrees sum to at least 19 and the three smallest sum to at least $\binom{3}{2} = 3$, or $A = 2, B = 5$, in which case the five largest degrees already sum to at least 22. Both contradict the total degree 21. Hence $\sum_i \Phi_\pi(z_i) = (A + B)/2 \leq 3$. The score sequence $(0, 1, 2, 3, 5, 5, 5)$ satisfies (6.21) and attains three. The deterministic values follow from Theorem 3.1. $\square$

6.4 Exact frontier for the expanded rule under on-sample deletion stability

The preceding laws are minimax over arbitrary invariant learners. When the augmented sample satisfies the event below, an explicit endpoint correction admits a substantially wider frontier. Define the leave-one reference fit

$$g_{-i} := \mathcal{A}\big((Z_r)_{r \neq i}\big)$$

and the label-invariant augmented-sample deletion-stability event

$$\mathcal{E}_\varepsilon := \left\{ \max_{i \neq k} \max_{1 \leq j \leq n} \left| \widehat{f}_{-\{i,k\}}(X^i_j) - g_{-i}(X^i_j) \right| \leq \varepsilon \right\}. \tag{6.27}$$

If $\mathcal{I}_{a,b}(x) = [L_{a,b}(x), U_{a,b}(x)]$, let

$$\mathcal{I}^{+2\varepsilon}_{a,b}(x) := [L_{a,b}(x) - 2\varepsilon, U_{a,b}(x) + 2\varepsilon]. \tag{6.28}$$

This is an explicit robustness correction, not an asymptotic stability assumption.

The next theorem compares every leave-two candidate with its leave-one reference and shows that endpoint inflation reduces failure to a scalar order-statistic event.

Theorem 6.5 (Exact frontier for the 2ε-expanded rule). *In the balanced model, fix $\varepsilon \geq 0$. For any $\gamma \in [0,1]$, if $1 \leq t \leq F$, $1 \leq \ell \leq m$, and $\mathbb{P}(\mathcal{E}_\varepsilon) \geq 1-\gamma$, then the expanded rule satisfies*

$$\mathbb{P}\{\textit{new block fails under } \mathcal{I}^{+2\varepsilon}_{a,b}\} \leq \frac{\ell}{N} + \frac{s-\ell}{N}\gamma, \qquad s := \min\{N, 2L+1\}, \tag{6.29}$$

where L is defined in (3.1)*. In particular, almost-sure on-sample deletion stability gives the sharp expanded-rule bound ℓ/N.*

For the full parameter range and every fixed $\varepsilon \geq 0$, the worst-case failure probability of the expanded rule over laws and learners satisfying $\mathcal{E}_\varepsilon$ almost surely is

$$\begin{cases} 0, & t = 0 \text{ or } \ell = 0, \\ 1, & 1 \leq \ell \leq m \text{ and } t > F, \\ \ell/N, & 1 \leq t \leq F, 1 \leq \ell \leq m. \end{cases} \tag{6.30}$$

Consequently, for $\delta \in (0,1)$ with $\lfloor \delta N \rfloor \geq 1$, the unique maximal finite-rank pair for the 2ε-expanded rule in the almost-sure on-sample stable class is

$$(t, \ell) = (F, \lfloor \delta N \rfloor), \tag{6.31}$$

and, along every sequence with $\min\{m,n\} \to \infty$ and fixed α, δ, the asymptotic validity region among fixed positive a, b is the rectangle

$$a \leq \alpha, \qquad b \leq \delta. \tag{6.32}$$

Over all $(a,b) \in [0,1)^2$, the full limiting region additionally contains the two degenerate rank-zero axes $\{a = 0\} \cup \{b = 0\}$. Every exact-law and frontier assertion in this theorem concerns the expanded rule; when $\varepsilon = 0$, it is the original endpoint rule.

Proof. On $\mathcal{E}_\varepsilon$, we first compare every candidate endpoint with a leave-one reference endpoint and obtain a common scalar interval contained in the expanded rule. A scalar order-statistic count controls this event, the general bound controls its complement, and a constant learner gives sharpness.

Define the reference radius

$$Q_i := q^+_{n,a}\left(\left(\left|Y^i_j - g_{-i}(X^i_j)\right|\right)^n_{j=1}\right). \tag{6.33}$$

On $\mathcal{E}_\varepsilon$, the one-Lipschitz property of order statistics gives $|R_{ik} - Q_i| \leq \varepsilon$. Each candidate endpoint for hypothetical test label i is therefore within 2ε of $g_{-i}(X^i_j) \pm Q_k$. Taking outer order statistics and then applying (6.28) shows that the expanded interval contains

$$\left[g_{-i}(X^i_j) - Q^{(-i)}_{(d)}, g_{-i}(X^i_j) + Q^{(-i)}_{(d)}\right], \qquad d = N - \ell, \tag{6.34}$$

where $Q^{(-i)}_{(d)}$ is the dth smallest of $(Q_k)_{k \neq i}$.

If the expanded interval misses at least F coordinates, the reference residual exceeds $Q^{(-i)}_{(d)}$ on at least F coordinates. Since $t \leq F$, this forces

$$Q_i > Q^{(-i)}_{(d)}. \tag{6.35}$$

Among N real numbers, at most ℓ indices satisfy (6.35); ties only reduce this number. Hence, pointwise on $\mathcal{E}_\varepsilon$, at most ℓ hypothetical labels fail the expanded rule. On its complement, expansion can only remove failures from the unexpanded rule, for which (4.8) gives at most s failures. Let $H_N^{+2\varepsilon}$ denote failure of the expanded rule. Exchangeability and the label invariance of $\mathcal{E}_\varepsilon$ give

$$\mathbb{P}\{H_N^{+2\varepsilon} = 1\} \leq \frac{1}{N}\big[\ell\mathbb{P}(\mathcal{E}_\varepsilon) + s\mathbb{P}(\mathcal{E}_\varepsilon^c)\big] = \frac{\ell}{N} + \frac{s-\ell}{N}\mathbb{P}(\mathcal{E}_\varepsilon^c).$$

Since $L \geq \ell - 1$, we have $s \geq \ell$; substituting $\mathbb{P}(\mathcal{E}_\varepsilon^c) \leq \gamma$ proves (6.29).

For sharpness under almost-sure stability, use the constant-zero learner, which is 0-stable. Give type i exactly F positive responses of magnitude q_i and set its other responses to zero, where the q_i's are strictly increasing with consecutive gaps larger than 2ε. When $1 \leq t \leq F$, its inner radius is q_i, and precisely the ℓ largest-radius types fail after the 2ε expansion. If $t > F$, and $1 \leq \ell \leq m$, give every type F responses of a common magnitude larger than 2ε and all remaining responses zero. Every inner radius is zero and every type fails. Uniformly permuting the types gives an exchangeable law; arbitrarily small distinct perturbations remove residual ties while preserving the gaps. The rank-zero cases follow from (2.8). This proves (6.30); the finite and asymptotic frontier statements follow by substitution. □

Remark 6.6 (Endpoint inflation is essential). For $\varepsilon > 0$, on-sample stability alone does not improve the minimax law of the unexpanded rule at positive finite ranks $1 \leq t \leq F$ and $1 \leq \ell \leq m$. Scale every prediction in the general sharpness construction by $c \in (0, \varepsilon/3]$ while keeping all responses zero. The leave-one reference fit is zero and every pairwise prediction differs from it by at most $3c \leq \varepsilon$, so $\mathcal{E}_\varepsilon$ holds. Positive scaling preserves all strict rank comparisons and failures, and the unexpanded rule still attains the corresponding general value $\min\{1, (2L+1)/N\}$. This argument does not apply at $\varepsilon = 0$, when the expanded and original rules coincide.

The mechanism parallels the stability analysis of ordinary jackknife+ in Barber et al. [1], but the target here is the probability that the realized covered fraction of an arbitrarily dependent test block falls below $1 - \alpha$. On $\mathcal{E}_\varepsilon$, every candidate endpoint is within 2ε of a reference endpoint, and the expanded aggregate contains the common reference interval in (6.34). The failure analysis therefore reduces to one scalar order-statistic comparison per block.

6.5 Approximate exchangeability via block swaps

The exact pointwise label count also degrades gracefully if block exchangeability is only approximate. This is an application of the swap-TV principle developed for conformal prediction beyond exchangeability by Barber et al. [2] and, in a general group-invariance form, by Dobriban and Yu [4].

Let P now be any law on the labeled N-block array, not necessarily exchangeable. For $i \leq N$, let τ_i swap labels i and N, with τ_N the identity, and define

$$\varepsilon_{\mathrm{swap}}(\mathsf{P}) := \frac{1}{N}\sum_{i=1}^{N} d_{\mathrm{TV}}\big(\mathsf{P}, (\tau_i)_{\#}\mathsf{P}\big), \tag{6.36}$$

where $d_{\mathrm{TV}}(P, Q) = \sup_A |P(A) - Q(A)|$.

Corollary 6.7 (Sharp block-swap degradation)**.** *In the balanced finite-rank setting, put* $s := \min\{N, 2(N-w)-1\}$. *Then*

$$\mathbb{P}_{\mathsf{P}}\{H_N = 1\} \leq \min\left\{1, \frac{s}{N} + \varepsilon_{\mathrm{swap}}(\mathsf{P})\right\}. \tag{6.37}$$

The coefficient one is sharp: whenever $s < N$, for every $0 \le \eta \le 1 - s/N$ there are a law P_η and an admissible learner for which

$$\varepsilon_{\mathrm{swap}}(\mathsf{P}_\eta) = \eta, \qquad \mathbb{P}_{\mathsf{P}_\eta}\{H_N = 1\} = \frac{s}{N} + \eta. \tag{6.38}$$

Proof. The upper bound averages the total-variation defect of each swap against the pointwise label count. Sharpness then tilts the finite extremal orbit toward the event that the designated test label fails.

Equivariance under the transposition gives

$$\mathbb{E}_{\mathsf{P}} H_i = \mathbb{E}_{(\tau_i)_{\#}\mathsf{P}} H_N.$$

The defining variational property of total variation therefore implies

$$\mathbb{E}_{\mathsf{P}} H_N \le \mathbb{E}_{\mathsf{P}} H_i + d_{\mathrm{TV}}\big(\mathsf{P}, (\tau_i)_{\#}\mathsf{P}\big).$$

Average this inequality over i and use the pointwise count $\sum_i H_i \le s$ to obtain (6.37).

For sharpness, let Q be the uniform permutation law of the extremal construction in Section 5, so exactly s hypothetical labels fail on every realization. Put $\gamma = s/N$, let $A = \{H_N = 1\}$, and tilt Q by

$$\frac{d\mathsf{P}_\eta}{d\mathsf{Q}} = \frac{\gamma + \eta}{\gamma}\mathbf{1}_A + \frac{1 - \gamma - \eta}{1 - \gamma}\mathbf{1}_{A^c}. \tag{6.39}$$

Then $\mathsf{P}_\eta(A) = \gamma + \eta$. Under Q, for every $i \ne N$, the indicators H_N, H_i are two draws without replacement from s ones and $N - s$ zeros. Consequently

$$\mathsf{Q}\{H_N \ne H_i\} = \frac{2\gamma(1-\gamma)N}{N-1},$$

while the two density levels in (6.39) differ by $\eta/[\gamma(1-\gamma)]$. Thus

$$d_{\mathrm{TV}}\big(\mathsf{P}_\eta, (\tau_i)_{\#}\mathsf{P}_\eta\big) = \frac{\eta N}{N-1} \quad (i \ne N),$$

and the identity term for $i = N$ is zero. Averaging gives $\varepsilon_{\mathrm{swap}}(\mathsf{P}_\eta) = \eta$, proving (6.38). $\square$

7 Consequences and efficiency

This section first instantiates the complete-failure phase at the conventional outer factor-two choice. It then gives an explicit pathwise length separation and a deterministic diagnostic for strict interval containment.

7.1 The naive outer correction can fail completely

Proposition 7.1 (Failure of the ordinary factor-two correction)**.** *Take*

$$m = 19, \quad n = 200, \quad \alpha = a = \delta = 0.2, \quad b = 0.1 = \delta/2.$$

Then

$$\Gamma_{19,200,0.2}(0.2, 0.1) = 1.$$

Proof. Here

$$N = 20, \quad F = 41, \quad t = 40, \quad \ell = 2,$$

and

$$L = \min\left\{19, \left\lfloor \frac{41}{2} \right\rfloor\right\} = 19.$$

The conclusion follows from Theorem 3.1. Thus an exchangeable law and a block-permutation-invariant learner exist for which every possible test type misses at least 41/200 responses. □

The phenomenon persists asymptotically by Corollary 3.4. The ordinary outer choice $b = \delta/2$ spends the full outer factor-two budget while leaving no budget for the inner coordinate fraction. As a smaller integer anchor, if $m = n = 19$ with the same four levels $\alpha = a = \delta = 0.2$, $b = 0.1$, then

$$F = t = 4, \quad \ell = 2, \quad L = 4, \qquad \Gamma = \frac{9}{20}.$$

7.2 An explicit atomless pathwise average-length separation

We next show, by an explicit existence construction, that a valid frontier point can be much shorter pathwise than jackknife-minmax.

Proposition 7.2 (Atomless pathwise length separation)**.** *Let*

$$m = n = 19, \qquad \alpha = \delta = 0.2, \qquad a = b = 0.1.$$

Then the endpoint rule is valid with

$$\Gamma_{19,19,0.2}(0.1, 0.1) = \frac{3}{20} < 0.2.$$

For the realized test block, define the pathwise average-length advantage

$$\Delta := \frac{1}{19} \sum_{j=1}^{19} \left[\mathrm{len}\big(\mathcal{I}_{\mathrm{mm}}(X_j^N)\big) - \mathrm{len}\big(\mathcal{I}_{a,b}(X_j^N)\big)\right]. \tag{7.1}$$

For every $K > 0$, there is an exchangeable block model on which all displayed intervals are finite and, for every realization of the type permutation,

$$\frac{1}{19}\sum_{j=1}^{19} \mathrm{len}(\mathcal{I}_{a,b}(X_j^N)) = 2, \qquad \frac{1}{19}\sum_{j=1}^{19} \mathrm{len}(\mathcal{I}_{\mathrm{mm}}(X_j^N)) = \frac{2K+38}{19}. \tag{7.2}$$

In particular, $\Delta = 2K/19$, while the endpoint-to-minmax average-length ratio is

$$\frac{38}{2K+38} \longrightarrow 0. \tag{7.3}$$

If the responses are replaced by independent continuous noise supported on $[-\xi, \xi]$, with $0 < \xi < K/38$, the law is atomless and, pathwise,

$$\Delta \geq \frac{2K - 76\xi}{19} > 0. \tag{7.4}$$

Moreover, the endpoint-rule average length is at most $2 + 2\xi$, while the minmax average length is at least $(2K + 38 - 38\xi)/19$. Their ratio is therefore at most

$$\frac{38(1+\xi)}{2K + 38 - 38\xi}, \tag{7.5}$$

which also tends to zero for fixed ξ.

This is an explicit-family existence separation. It does not assert expected-length minimax optimality or uniform dominance over a natural model class.

Proof. We first verify that the selected ranks lie on the exact validity frontier. A cyclic prediction table then gives the two pathwise lengths, after which a bounded atomless response perturbation establishes the atomless claim.

The ranks are

$$N = 20, \quad F = 4, \quad t = 2, \quad \ell = 2, \quad D_\delta = 4, \quad q = 2.$$

The criterion (3.9) holds because

$$F(q - \ell + 1) = 4 \geq 3 = q(t-1) + 1.$$

Moreover $L = \lfloor 4/3 \rfloor = 1$, giving the stated worst-case failure probability. In addition,

$$T_2 = 1 + \left\lfloor \frac{4(2-2+1)-1}{2} \right\rfloor = 2,$$

so $(t, \ell) = (2, 2)$ is a maximal point of the exact lattice Pareto antichain, not merely an interior valid pair.

For the length comparison, label the 20 block types by $u \in \mathbb{Z}/20\mathbb{Z}$, give type u covariates (u, j), $j = 1, \ldots, 19$, and uniformly permute the types. Define one learner by first specifying the following pairwise table. When the training multiset omits $\{u, v\}$, its prediction at the first coordinate of type u is

$$\widehat{f}_{-\{u,v\}}(u, 1) = \begin{cases} -K, & v - u \equiv 1 \pmod{20}, \\ K, & v - u \equiv -1 \pmod{20}, \\ 0, & \text{otherwise.} \end{cases} \tag{7.6}$$

As $v \neq u$ varies, these values are

$$-K, \ \underbrace{0, \ldots, 0}_{17 \text{ times}}, \ K.$$

At coordinates $j = 2, \ldots, 19$ of either omitted type, set every prediction to 1. The additive type-count formula (5.3) realizes this table and is measurable and block-permutation-invariant. Initially set all responses to zero. Each calibration block has at most one residual different from 1, so both the $a = 0.1$ score of rank 18 and the $\alpha = 0.2$ score of rank 16 equal 1.

At coordinate 1, the minmax interval and the endpoint interval are

$$[-K-1, K+1] \quad \text{and} \quad [-1, 1],$$

with lengths $2K + 2$ and 2. At each other coordinate both rules return $[0, 2]$. This proves (7.2), (7.3), and the stated difference.

Now let every response be an independent continuous random variable supported on $[-\xi, \xi]$, and then uniformly permute the noisy types. The learner ignores the responses and still reads only the unordered covariate-labelled training blocks. Conditional on the discrete type permutation, the response vector has a product atomless law, so the resulting exchangeable law is atomless.

Relative to the zero-response construction, every residual moves by at most ξ. The one-Lipschitz property of order statistics therefore moves each relevant inner score by at most ξ. Consequently each outer candidate endpoint, each outer endpoint quantile, and Q_δ moves by at most ξ; the length of either interval changes by at most 2ξ. At coordinate 1, the minmax length minus the endpoint length is thus at least $2K - 4\xi$. At each of the other 18 coordinates it is at least -4ξ.

Averaging gives (7.4). The bound is positive when $\xi < K/38$, and it holds for every realization in the support, not only in expectation. The preceding Lipschitz bound also shows that every endpoint-rule coordinate has length at most $2 + 2\xi$. Moreover, $Q_\delta \geq 1 - \xi$, so the minmax average length is at least $(2K + 38 - 38\xi)/19$. Dividing proves (7.5). $\square$

7.3 A general strict-containment diagnostic

The preceding example isolates a general deterministic condition under which endpoint trimming strictly contracts the minmax interval.

Proposition 7.3 (A sufficient condition for strict containment)**.** *Fix a covariate x, write*

$$p_i := \widehat{f}_{-i}(x), \qquad M_a := \max_i S_i(a), \qquad Q_\delta := q^+_{m,\delta}\big((S_i(\alpha))_{i=1}^m\big),$$

and let $p_{(1)} \leq \cdots \leq p_{(m)}$. Assume the endpoint interval is nonempty, $1 \leq \ell \leq m$, and all the displayed predictions and score radii are finite. If

$$p_{(\ell)} - p_{(1)} > M_a - Q_\delta, \qquad p_{(m)} - p_{(N-\ell)} > M_a - Q_\delta, \tag{7.7}$$

then

$$\mathcal{I}_{a,b}(x) \subset \operatorname{int} \mathcal{I}_{\mathrm{mm}}(x).$$

Proof. Coordinatewise monotonicity gives

$$q^-_{m,b}\big((p_i - S_i(a))_i\big) \geq p_{(\ell)} - M_a > p_{(1)} - Q_\delta.$$

Similarly,

$$q^+_{m,b}\big((p_i + S_i(a))_i\big) \leq p_{(N-\ell)} + M_a < p_{(m)} + Q_\delta.$$

These are precisely the two strict endpoint inclusions. Because the inequalities in (7.7) are strict, they describe an open class of finite endpoint and score arrays. The proposition is a pointwise diagnostic, not a claim of uniform expected-length dominance. $\square$

Operationally, the condition is favored when $M_a - Q_\delta$ is small relative to both prediction gaps in (7.7). This occurs, for example, when score radii are controlled and nearly homogeneous while a small number of leave-environment predictions form separated tails.

8 Meaning, related work, and boundaries

We first interpret the two terms in the rank budget and the role of learner stability. We then locate the result relative to conformal, hierarchical, and batch prediction, before recording the remaining scope boundaries.

8.1 What the exact law explains

The asymptotic budget (3.16) separates two distinct costs. The term $2b/\delta$ is the familiar outer jackknife+ tournament penalty. The new term a/α records the inner allocation needed to control a fraction of coordinates in the test block. Different coordinates can be missed by different candidate environments, so a one-response jackknife+ argument cannot simply be repeated coordinatewise.

The complete-failure region (3.17) shows that this is a phase transition, not a small undercoverage effect. Spending the entire outer factor-two budget can make every possible test label fail. The heterogeneous theorem gives the same explanation vertex by vertex: each block size induces an outdegree demand, and simultaneous failure is possible exactly when the incident-edge system can meet those demands. The asymmetric theorem then shows that assigning different budgets to the two tails cannot evade the obstruction.

The stability theorem identifies what is worst-case about that picture. On $\mathcal{E}_\varepsilon$, every candidate endpoint is within 2ε of a reference endpoint, and the 2ε-expanded aggregate interval contains the common reference interval used in the scalar comparison. When $1 \le t \le F$ and $1 \le \ell \le m$, the expanded rule then has exact worst-case failure probability ℓ/N over the almost-sure $\mathcal{E}_\varepsilon$ class, and it has the rectangular limiting region (6.32). For $\varepsilon > 0$, Remark 6.6 shows that this conclusion does not extend to the unexpanded rule. Thus the general frontier reflects the cost of allowing test coordinates to interact with an unrestricted leave-environment learner.

The frontier is not solely an impossibility boundary. The explicit family in Proposition 7.2 removes two separated prediction tails while keeping its pathwise average length bounded; the jackknife-minmax average length diverges, and the separation survives bounded atomless response perturbations. This is an existence separation, not a uniform efficiency claim.

8.2 Relation to prior work

Conformal and leave-one-out prediction. Distribution-free predictive inference originates with conformal prediction [18]; cross-conformal prediction combines foldwise conformal scores from cross-fitted models [17]. The ordinary jackknife+ theorem of Barber et al. [1] proves $1 - 2b$ coverage, establishes sharpness of the factor two, and recovers a $1 - b$ guarantee with jackknife-minmax. Nested conformal and quantile out-of-bag methods provide a broader endpoint-based perspective [9]. We do not claim the comparison tournament itself as new. The new combinatorial step is the coordinate–environment incidence layer, together with an exact realization showing that no information is lost in passing from block failures to degree demands. Our stable-learner result is likewise a block-level exact analogue of the stability mechanism in ordinary jackknife+, under the simultaneous on-sample event (6.27) and endpoint expansion.

Hierarchical and multi-environment inference. The leave-two-environment augmentation follows the proof architecture of Theorem 1 of Duchi et al. [5]. Their model permits random unequal block sizes and arbitrary conditional block laws; their extreme envelope avoids an endpoint-rank loss. Duchi et al. do not provide a finite-sample validity theorem for their Algorithm 9 and report empirical undercoverage in Appendix D. Theorem 6.1 gives the exact conditional minimax envelope for the class with a fixed realized unordered size multiset, positive induced inner ranks, and a positive outer rank, together with the conditional and expectation bounds for random sizes.

For different targets, Dunn et al. [6] study a hierarchical-i.i.d. model and predict one response from a new subject by pooling empirical distribution functions and by subsampling. Lee et al. [13] formulate hierarchical exchangeability, allowing unequal random group sizes but requiring within-group exchangeability; their hierarchical jackknife+ gives marginal coverage for one response, while their stronger repeated-measurement guarantee uses conditionally i.i.d. repeats. SymmPI [4] covers general group invariances and random imbalanced sizes, but its hierarchical specialization again assumes within-group exchangeability and targets a missing response. None of these results controls the realized covered fraction of a new block under arbitrary within-block dependence.

Exact batch coverage has a separate and closely related literature. Gazin et al. [8] derive the joint law of multiple conformal ranks, and Marques F. [15] gives the exact empirical-coverage distribution for split conformal prediction. Batch predictive inference [14] directly controls the empirical covered fraction of an exchangeable test batch and proves fixed-rank optimality from an exact test-rank law.

Those results assume individual-level exchangeability of the calibration and test observations and scores not produced by coupled leave-two-environment fits. Here the exchangeable units are whole blocks, coordinates within a block may be arbitrarily dependent, and each endpoint is built from mutually coupled leave-environment predictions. The incidence and oriented-graph obstruction is absent from the individual-rank batch model. Cross-validation conformal risk control [3] also targets a batch loss, using a minmax-type union rather than the endpoint frontier studied here. Quantiles of local quantiles also appear in one-shot federated conformal prediction [11], but there the target is one test response and validity comes from exchangeable individual scores rather than coupled leave-environment fits.

Beyond exchangeability and graph realization. Swap total-variation bounds for nonexchangeable conformal and jackknife+ procedures were developed by Barber et al. [2]; SymmPI gives a general transformation-group formulation. Corollary 6.7 specializes that principle to whole-block swaps around our sharp pointwise baseline and proves that its coefficient is attained by the same extremal orbit.

The remaining combinatorial tools are classical. Landau's theorem characterizes tournament scores [12]; Hall's theorem supplies the incident-edge matching for heterogeneous demands [10]; and the Gale–Ryser theorem realizes equal-row-sum binary matrices. Our claim is therefore deliberately specific: to our knowledge, prior work does not give the exact finite-sample minimax failure law, the fixed-multiset heterogeneous degree-demand frontier with positive ranks, or the hidden-global-rank tournament optimization for balanced endpoint-quantile multi-environment jackknife+ under arbitrary within-block dependence. This is not a priority claim for hierarchical conformal prediction or tournament arguments in general.

8.3 Scope and open directions

The balanced theorem covers every deterministic rank cell, while the unequal size theorem assumes $t_i \geq 1$ for every realized size type. Mixed zero and positive inner ranks create infinite radii on only some vertices, coupling candidate availability across the graph; their exact frontier is open. The random-size upper bound (6.12) remains valid whenever the positive-rank condition holds almost surely.

Randomized tie-breaking cannot repair an invalid deterministic cell because Proposition 5.3 gives tie-free extremizers. Theorem 6.3 instead analyzes one global, data-independent random outer rank with fixed inner rank. Joint randomization of both levels, data-dependent selection of a Pareto point, and procedures in which the learner observes the rank coin lead to different minimax games. Valid and efficient adaptive selection among the finite frontier points is a natural next question.

The on-sample stability condition (6.27) is simultaneous over all augmented deletions and observed covariates. Weaker in-probability, average, or distribution-dependent notions may interpolate between (6.29) and the general exact law more favorably. Within-block i.i.d. sampling and smooth hierarchical models may also permit shorter intervals, but they define restricted distribution classes. Optimizing expected length over such a class, rather than exhibiting a strict pathwise gain, remains open.

Finally, a coherently randomized learner is covered when the same independent seed U is used in every augmented fit. Independently rerandomizing the same training multiset destroys the single comparison graph. The same issue arises if the learner observes the global random rank in Theorem 6.3; in either case a different coupling analysis is needed.

9 Conclusion

Endpoint quantiles can replace the extreme multi-environment prediction envelope, but their two ranks share an exact finite-sample budget. The balanced law, its lattice frontier, and the matching graph construction show that the factor-two outer penalty and the new inner incidence penalty are both structural. Hall feasibility extends the result to fixed unequal sizes with positive inner and outer ranks. In the balanced model, asymmetric ranks, a hidden global random rank, the expanded rule under on-sample deletion stability, and finite-rank approximate exchangeability identify which parts of the obstruction persist under altered assumptions.

References

[1] R. F. Barber, E. J. Candès, A. Ramdas, and R. J. Tibshirani. Predictive inference with the jackknife+. *The Annals of Statistics*, 49(1):486–507, 2021. doi:10.1214/20-AOS1965. https://arxiv.org/abs/1905.02928.

[2] R. F. Barber, E. J. Candès, A. Ramdas, and R. J. Tibshirani. Conformal prediction beyond exchangeability. *The Annals of Statistics*, 51(2):816–845, 2023. doi:10.1214/23-AOS2276. https://arxiv.org/abs/2202.13415.

[3] K. M. Cohen, S. Park, O. Simeone, and S. Shamai (Shitz). Cross-validation conformal risk control. In *Proceedings of the 2024 IEEE International Symposium on Information Theory*, pages 250–255, 2024. doi:10.1109/ISIT57864.2024.10619172. https://arxiv.org/abs/2401.11974.

[4] E. Dobriban and M. Yu. SymmPI: Predictive inference for data with group symmetries. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 87(5):1353–1381, 2025. doi:10.1093/jrsssb/qkaf022. https://arxiv.org/abs/2312.16160.

[5] J. C. Duchi, S. Gupta, K. Jiang, and P. Sur. Predictive inference in multi-environment scenarios. *Statistical Science*, 40(3):392–416, 2025. doi:10.1214/24-STS973. https://arxiv.org/abs/2403.16336v2.

[6] R. Dunn, L. Wasserman, and A. Ramdas. Distribution-free prediction sets for two-layer hierarchical models. *Journal of the American Statistical Association*, 118(544):2491–2502, 2023. doi:10.1080/01621459.2022.2060112. https://arxiv.org/abs/1809.07441.

[7] D. Gale. A theorem on flows in networks. *Pacific Journal of Mathematics*, 7:1073–1082, 1957. doi:10.2140/pjm.1957.7.1073.

[8] U. Gazin, G. Blanchard, and E. Roquain. Transductive conformal inference with adaptive scores. In *Proceedings of the 27th International Conference on Artificial Intelligence and Statistics*, volume 238 of *Proceedings of Machine Learning Research*, pages 1504–1512, 2024. https://proceedings.mlr.press/v238/gazin24a.html.

[9] C. Gupta, A. K. Kuchibhotla, and A. Ramdas. Nested conformal prediction and quantile out-of-bag ensemble methods. *Pattern Recognition*, 127:108496, 2022. doi:10.1016/j.patcog.2021.108496.

[10] P. Hall. On representatives of subsets. *Journal of the London Mathematical Society*, s1-10(1):26–30, 1935. doi:10.1112/jlms/s1-10.37.26.

[11] P. Humbert, B. Le Bars, A. Bellet, and S. Arlot. One-shot federated conformal prediction. In *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pages 14153–14177, 2023. https://proceedings.mlr.press/v202/humbert23a.html.

[12] H. G. Landau. On dominance relations and the structure of animal societies: III. The condition for a score structure. *Bulletin of Mathematical Biophysics*, 15:143–148, 1953. doi:10.1007/BF02476378.

[13] Y. Lee, R. F. Barber, and R. Willett. Distribution-free inference with hierarchical data. *ACM Journal of Data Science*, 2026. doi:10.1145/3786352. https://arxiv.org/abs/2306.06342.

[14] Y. Lee, E. Tchetgen Tchetgen, and E. Dobriban. Batch predictive inference. arXiv:2409.13990v5, 2025. https://arxiv.org/abs/2409.13990v5.

[15] P. C. Marques F. Universal distribution of the empirical coverage in split conformal prediction. *Statistics & Probability Letters*, 219:110350, 2025. doi:10.1016/j.spl.2024.110350. https://arxiv.org/abs/2303.02770.

[16] H. J. Ryser. Combinatorial properties of matrices of zeros and ones. *Canadian Journal of Mathematics*, 9:371–377, 1957. doi:10.4153/CJM-1957-044-3.

[17] V. Vovk. Cross-conformal predictors. *Annals of Mathematics and Artificial Intelligence*, 74(1–2):9–28, 2015. doi:10.1007/s10472-013-9368-4.

[18] V. Vovk, A. Gammerman, and G. Shafer. *Algorithmic Learning in a Random World*. Springer, 2005. doi:10.1007/b106715.