FROM PER-CLASS SUMMARIES TO TOP-k DECISIONS: SHARP INFORMATION LIMITS

Anonymous authors
Paper under double-blind review

ABSTRACT

When deploying a classifier ahead of an expensive reranker or human expert, practitioners must choose a shortlist size k that captures the true label while limiting downstream costs. Can this shortlist size be reliably chosen from standard evaluation reports containing top-1 accuracy and per-class score distributions?

We show that these reports need not suffice: two evaluation sets can share identical top-1 accuracy and per-class summaries, yet yield substantially different top-k coverages. This ambiguity arises when per-class summaries do not determine how scores across different classes co-occur on individual examples.

To understand the fundamental limits of this information loss, we first analyze the idealized setting of true posterior probabilities. Even with ideal scores, the maximum coverage gap between report-compatible distributions is $\frac{1}{5}$ at top-2 and $\frac{17}{76}$ at top-3, converging to $\frac{1}{4}$ as both shortlist size and class count grow. We establish these sharp limits using dual certificates and matching constructive examples, which provide both optimistic and guaranteed shortlist bounds for any supplied report.

For finite-support uncalibrated classifiers, exact programs certify actual coverage and shortlist requirements. Across 150 benchmark reports with native five-vote or five-neighbor scores, required shortlist sizes are uniquely determined at 90% target coverage, but 29 reports remain ambiguous at 95%—in one case swinging the required shortlist from two to five candidates. Adding joint cross-class statistics narrows this uncertainty, while a compact true-label rank histogram eliminates it entirely. Our results pinpoint when standard per-class summaries suffice for shortlist sizing and what additional statistics evaluators should report.

1 INTRODUCTION

A classifier can serve as the first stage of a larger decision pipeline, passing a shortlist of candidate labels to a more expensive model or human reviewer. In domain classification, for example, a lightweight shortlister selects candidates for a richer contextual reranker (Kim et al., 2018). Larger shortlists require processing more candidates; smaller shortlists risk discarding the correct label before the second stage sees it. If the second stage must choose from the shortlist, its accuracy cannot exceed the first stage's true-label coverage. The user therefore needs the smallest shortlist that meets a desired coverage target.

Now consider a user making this decision from an evaluation report, as when reusing benchmark results or handing a model from one team to another. The evaluator supplies top-1 accuracy and per-class diagnostics without explicit links joining scores across classes on individual examples. Can this report justify passing two candidates to the next stage, or might five be required to retain the correct label 95% of the time?

We show that even complete per-class summaries can leave this decision unresolved. Two evaluation batches can display identical top-1 accuracy and identical per-class score and label distributions, yet require different shortlist budgets to reach the same coverage target. Figure 1 illustrates this phenomenon on a standard digit classification benchmark: two integer batches share the exact same top-1 accuracy (91.67%) and the same per-class diagnostic marginals, yet one batch achieves 95% true-label coverage with 2 candidates, while the other requires 5. This gap changes how many candidates must reach the second stage. It arises from how class scores *co-occur within the same example*, an association not determined by this report.

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

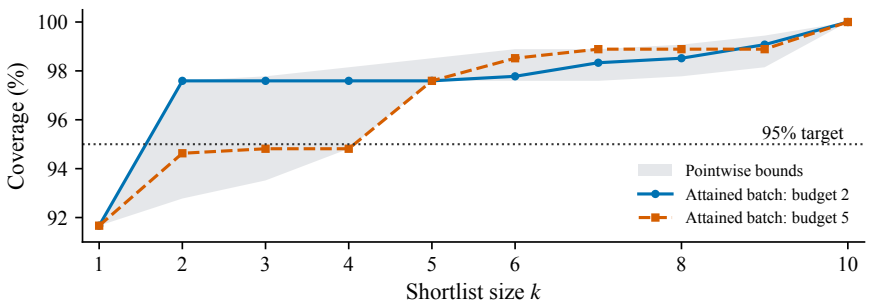


Figure 1: **Identical per-class summaries permit divergent shortlist sizing.** Two compatible evaluation batches for an illustrative Digits/Random Forest report share identical top-1 accuracy (91.67%) and matching per-class score/label distributions, yet require 2 versus 5 candidates to reach 95% true-label coverage on their respective batches. The shaded envelope marks point-by-point coverage limits $[L_k, U_k]$, while solid curves show realized performance for two concrete, complete integer batches. See Section 5 and Section I for full details.

We study a deliberately rich per-class report: for every class $i \in \{1, \ldots, N\}$, the joint distribution of its assigned score and one-vs-rest binary label, along with overall top-1 accuracy. This includes every threshold-based one-vs-rest ROC, precision–recall, and classwise calibration curve, so reporting finer single-class diagnostics cannot remove any remaining ambiguity. Our question is decision-specific: *When does the report determine the shortlist budget, and what additional summary should the evaluator retain when it does not?*

Sharp information limits in the continuum. To establish the fundamental boundaries of this ambiguity, we first analyze the idealized setting of true posterior probabilities (or vector-calibrated scores). Here, one might expect that knowing every 1-D posterior distribution and top-1 accuracy severely restricts ranking variability. Indeed, arbitrary column permutations are heavily constrained: they must preserve both simplex normalization ($\sum_i q_i = 1$) and top-1 accuracy ($\mathbb{E}[\max_i q_i]$). Nevertheless, we prove that nonidentification holds for every non-trivial rank $2 \le k < N$ in all finite dimensions $N \ge 3$.

We establish the *exact worst-case coverage gaps* between report-compatible distributions. By constructing explicit separable dual certificates and pairing them with matching extremal families, we show that the dimension-uniform supremum of the coverage gap is precisely $E_2 = 1/5$ (20.00 percentage points at top-2) and $E_3 = 17/76$ (≈ 22.37 percentage points at top-3), converging to $E_\infty = 1/4$ (25.00 percentage points) as both the shortlist rank k and class count N grow (Section 4). These are universal limits: no two posterior distributions consistent with the same per-class report can ever differ in top-2 coverage by more than $1/5$, and this bound cannot be improved.

From coverage intervals to certified budget decisions. For shortlist sizing, the key question is whether coverage uncertainty changes the minimum budget at the chosen target. We characterize exactly when this occurs (corollary 4.3). For any user-specified coverage target $\tau \in (0, 1]$, a report determines the optimistic minimum shortlist $b_{\text{possible}}(\tau)$ that *could* meet the target, and the certified minimum shortlist $b_{\text{certified}}(\tau)$ *guaranteed* to do so under all report-compatible distributions. A candidate budget is uniquely identified if and only if no rank-wise coverage interval straddles the target τ.

What a user can infer from 150 benchmark reports. Moving beyond idealized posteriors to uncalibrated classifiers, we solve the exact coverage programs over finite score supports for 150 reports that we construct from benchmark evaluations (5 datasets, 3 model families, and 10 splits; Section 5). We find that: (i) per-class reports reduce average top-2 coverage width by 51%–85% relative to the top-1 baseline $[A_1, 1]$ across non-degenerate dataset/model cells; (ii) at a target of $\tau = 90\%$, budgets are uniquely identified for all 150 reports, including reports requiring multiple candidates; (iii) at higher stringency ($\tau = 95\%$), however, 29 reports exhibit ambiguous budgets (swinging by up to 3 candidates), rising to 85 ambiguous reports at $\tau = 99\%$; and (iv) incorporating richer cross-class summaries (such as the argmax confusion matrix or maximum score distributions)

tightens the bounds in our illustrative report. A true-label rank histogram suffices to resolve all shortlist decisions without releasing individual evaluation records.

The reporting implication is concrete: retain information about the decisions the next stage must make. Complete classwise diagnostics may justify a shortlist budget at one target and leave it unresolved at another; a compact summary of true-label ranks supplies the missing information for every target.

2 Related Work

Shortlist coverage and decision-aware evaluation. Shortlisting controls the candidates processed by a subsequent stage. Kim et al. (2018) combine lightweight domain shortlisting with contextual reranking; Li et al. (2022) calibrate candidate pruning to control ranking loss while reducing reranking cost. Conformal risk control also studies set-valued predictions using calibration data (Angelopoulos et al., 2024). In multi-class classification and information retrieval, candidate list selection is typically analyzed through top-k surrogate consistency (Yang & Koyejo, 2020), generalized gain functions (Alvim et al., 2012), multiple guessing bounds (Kurri et al., 2021), or list-error formulations (Asadi Kangarshahi & Guillén i Fàbregas, 2023; Sason & Verdú, 2018). In particular, Sason & Verdú (2018) bound list-decision error using Arimoto–Rényi conditional entropy. We ask what can be inferred when only an evaluation report is available: the exact range of top-k coverage among laws sharing every classwise score/label distribution and top-1 accuracy, and the resulting range of candidate budgets. These are report-conditioned identification bounds, rather than finite-sample confidence guarantees.

Report reuse and statistical summaries. Evaluation tools support configurable metrics, sliced dashboards, and compact statistical profiles (TensorFlow, 2024; 2022; WhyLabs, n.d.). We study a particular export containing complete classwise diagnostics and top-1 accuracy; tools can also retain top-k metrics or other joint summaries. Evaluating whether such summary representations preserve downstream utility connects broadly to usable information under computation or evaluation constraints (Xu et al., 2020; Dubois et al., 2020) and task-specific representation games (Uzan & Weinberger, 2024; Zhao et al., 2022). Similarly, Jaeger et al. (2023) emphasize how surrogate evaluation protocols can obscure downstream failure modes. Our report-reuse question concerns the handoff from an evaluator to a user choosing a shortlist budget. The enrichment study makes this a reporting-design question: which additional aggregates resolve the user's decision?

Calibration: vector vs. classwise. Calibration is central to interpreting multi-class confidence scores (Guo et al., 2017; Chidambaram & Ge, 2025; Futami & Nitanda, 2026). Recent work emphasizes decision-aware calibration and reconciliation across different scoring heads (Du et al., 2025; Tang et al., 2026). In our analysis, we maintain a crucial distinction between true posterior probabilities (or full vector calibration, $\Pr\{X = i \mid S\} = S_i$) and classwise calibration ($\Pr\{X = i \mid S_i\} = S_i$). While true posteriors admit universal information limits ($E_2 = 1/5$, $E_3 = 17/76$), we show that classwise calibration alone does not equate expected score sums with actual label coverage (Section A.1). This distinction ensures that universal posterior bounds are not conflated with arbitrary uncalibrated scores, which we handle separately through exact linear programs.

Constrained coupling and extremal characterizations. Mathematically, bounding top-k sums across marginal distributions relates to complete mixability and multi-marginal coupling (Puccetti et al., 2012; Koike et al., 2024), threshold representations of order statistics (Ogryczak & Tamir, 2003), and rank-marginal bounds (Rychlik, 1993; 1994; Jaworski & Rychlik, 2008). Classical rank-marginal relaxations do not enforce simplex normalization row-by-row. On our two worked reports, this permits overly low lower endpoints and wider intervals; the relaxed upper endpoints are attainable by posterior laws (Section G.3). Variational formulations with linear constraints date back to Kellerer (1984) and Zaev (2015). Our theoretical contribution lies in carrying out the explicit, simplex-constrained extremal calculation: establishing sharp separable dual certificates at ranks two and three, matching constructive distributions, and providing unified budget guarantees.

3 REPORTS, COMPATIBILITY, AND COVERAGE

Consider a multi-class prediction task with label space $[N] = \{1, \dots, N\}$. Let $X \in [N]$ denote the ground-truth class, and let $S \in \Delta_N := \{s \in [0,1]^N : \sum_{i=1}^N s_i = 1\}$ be a predicted probability vector. When scores tie, we break ties deterministically by ascending class index. The first stage retains the top-k candidate classes for subsequent processing. We study a fixed candidate set with one true label per example. The shortlist succeeds whenever it contains that label, yielding the actual top-k coverage

$$A_k = \Pr\big\{X \in \mathrm{Top}_k(S)\big\}. \tag{1}$$

We consider a recipient who receives a summary instead of the joined per-example evaluation table $(s^{(m)}, x^{(m)})_{m=1}^M$. Formally, a *complete classwise evaluation report* is

$$\rho = \big(N, (\gamma_i)_{i=1}^N, a_1\big), \qquad \gamma_i = \mathrm{Law}\big(S_i, \mathbf{1}\{X = i\}\big), \quad a_1 = A_1. \tag{2}$$

This report retains the full joint distribution of score and binary label for each individual class (from which threshold-based one-vs-rest ROC and precision–recall curves, as well as classwise calibration curves, can be reconstructed; proposition A.1), together with overall top-1 accuracy. This report omits explicit row links, but exact empirical atoms can recover them through normalization (proposition G.1). Population marginals need not identify coverage. Our empirical programs concern finite-resolution scoring rules, rather than binned reports retaining the original continuous-score ranking.

The posterior baseline. To establish the fundamental theoretical limits of this reporting format, we first analyze the best-case idealized scenario: true posterior probabilities. Let Z be the input feature, and let $Q \in \Delta_N$ with $Q_i = \Pr\{X = i \mid Z\}$ be the true Bayes posterior vector distributed according to law $\mu \in \mathcal{P}(\Delta_N)$. For any realization $q \in \Delta_N$, sort coordinates in descending order $q_{(1)} \geq q_{(2)} \geq \cdots \geq q_{(N)}$. The optimal k-candidate shortlist simply selects the k classes with largest posterior mass, giving conditional coverage $F_k(q) = \sum_{j=1}^k q_{(j)}$. The expected top-$k$ coverage under μ is therefore

$$R_k(\mu) = \mathbb{E}_\mu\big[F_k(Q)\big]. \tag{3}$$

Because $\Pr\{X = i \mid Q\} = Q_i$, this Bayes risk $R_k(\mu)$ coincides exactly with the actual coverage A_k. The same identity holds for learned scores S if they are *full-vector calibrated* ($\Pr\{X = i \mid S\} = S_i$). Classwise calibration alone is not sufficient to guarantee this equivalence (Section A.1).

What a posterior report fixes. Let $\xi_{\mu,i} = \mathrm{Law}_\mu(Q_i)$ denote the 1-D marginal distribution of class i's posterior score, and let $\boldsymbol{\xi}_\mu = (\xi_{\mu,1}, \dots, \xi_{\mu,N})$. Under the Bayes posterior condition, the joint score/label law γ_i and the marginal $\xi_{\mu,i}$ determine each other via the identity

$$\Pr\{Q_i \in B, X = i\} = \int_B s\, \xi_{\mu,i}(\,\mathrm{d}s), \qquad \Pr\{Q_i \in B, X \neq i\} = \int_B (1-s)\, \xi_{\mu,i}(\,\mathrm{d}s). \tag{4}$$

These marginals also fix the class priors $p_i = \mathbb{E}[Q_i] = \int_0^1 x\, \xi_{\mu,i}(\,\mathrm{d}x)$. Conversely, any probability measure μ on the simplex Δ_N can be realized by a valid joint distribution (Q, X) by first sampling $Q \sim \mu$ and then sampling $X \mid Q \sim Q$ (Section A).

Compatible laws and coverage intervals. Given a feasible posterior summary $s = (N, \boldsymbol{\xi}, r)$, the recipient only knows that the data-generating law belongs to the compatibility set

$$\mathcal{C}(s) = \big\{\mu \in \mathcal{P}(\Delta_N) : \boldsymbol{\xi}_\mu = \boldsymbol{\xi},\ R_1(\mu) = r\big\} \neq \varnothing. \tag{5}$$

Because the report does not record joint cross-class couplings, different compatible laws in $\mathcal{C}(s)$ can yield different top-k coverages. We define the achievable coverage endpoints and the resulting ambiguity width as

$$L_k(s) = \min_{\mu \in \mathcal{C}(s)} R_k(\mu), \qquad U_k(s) = \max_{\mu \in \mathcal{C}(s)} R_k(\mu), \qquad W_k(s) = U_k(s) - L_k(s). \tag{6}$$

Both extrema are attained by valid measures (proposition 4.2); for any $k \geq N$, trivially $L_k(s) = U_k(s) = 1$. Table 1 outlines how this formulation connects idealized posteriors to empirical classifiers and finite batches.

Table 1: **Overview of reporting frameworks and their operational guarantees.** All bounds condition directly on the supplied report without asserting finite-sample population confidence intervals. The posterior setting yields universal theoretical limits, while empirical evaluation relies on report-specific linear programs and integer programming.

Setting	Coverage target	Guarantee
Posterior / vector-calibrated	$R_k = \mathbb{E}F_k$, equal to hits	Sharp universal bounds; attainable intervals
Arbitrary learned scores	Actual label hits A_k	Report-specific weighted-law intervals
Fixed-size batches	Integer label-hit counts	Attainable integer extrema; may differ from weighted laws

4 Sharp Information Limits and Budget Certificates

How much uncertainty can a per-class report induce in downstream ranking? To answer this, we study the worst-case coverage ambiguity over all possible data-generating posterior laws and all feasible reports. Formally, define the global posterior ambiguity at a fixed shortlist size k as

$$E_k = \sup_{N>k} \sup_{s \text{ feasible}} W_k(s). \tag{7}$$

In this supremum, both the label dimension N and the report s vary, but every competing pair of distributions shares the exact same report $(N, \boldsymbol{\xi}, r)$.

Theorem 4.1 (General-rank ambiguity)**.** ***(i) Universal nonidentification.*** *For every finite class count $N \geq 3$ and every nontrivial rank $2 \leq k < N$, there exist discrete posterior distributions that share identical labeled marginals, identical top-1 accuracy, and a common uniform prior, yet produce strictly different top-k coverages (R_k).*

(ii) Dimension-uniform bounds. *For every shortlist size $k \geq 2$, the global ambiguity satisfies*

$$\frac{1}{4} - \frac{1}{2(k^2+4k-2)} \leq E_k \leq \frac{1}{4}, \tag{8}$$

with the exact sharp values at ranks two and three given by

$$E_2 = \frac{1}{5} = 0.20, \qquad E_3 = \frac{17}{76} \approx 0.2237. \tag{9}$$

(iii) Large-rank asymptotic limit. *As the shortlist budget grows ($k \to \infty$), allowing the label space dimension N to grow as well, the worst-case ambiguity converges:*

$$\lim_{k\to\infty} E_k = \frac{1}{4}. \tag{10}$$

The upper bounds hold for all Borel probability measures on the simplex. The lower bounds are matched constructively by sequences of finite discrete distributions.

Crucially, theorem 4.1 demonstrates that even under the most generous statistical conditions—calibrated posteriors, infinite sample size, and every single-class diagnostic preserved—dropping cross-class co-occurrences creates an irreducible coverage ambiguity of up to 20 percentage points at rank two and nearly a quarter at larger ranks. At ten classes, the worst-case posterior top-two ambiguity lies between $7/48$ and $7/45$, or 14.58–15.56 percentage points (Sections F.2 and F.3).

Pooled marginals and what the report fixes. To derive upper bounds on ambiguity, we observe that pooling all N coordinate marginals yields an aggregate score measure $\nu_\mu = \sum_{i=1}^{N} \xi_{\mu,i}$ on $[0,1]$, with total mass N and first moment $\int_0^1 x\, \nu_\mu(\mathrm{d}x) = 1$. For any continuous test function $\psi \in C([0,1])$, the report uniquely determines the symmetric sum:

$$\mathbb{E}_\mu\left[\sum_{i=1}^{N} \psi(Q_i)\right] = \int_0^1 \psi(x)\, \nu_\mu(\mathrm{d}x). \tag{11}$$

Any valid upper bound certified through ν_μ holds automatically for the richer class-indexed report. Conversely, our matching lower-bound constructions preserve every individual class marginal $\xi_{\mu,i}$ identically.

A universal separable certificate for arbitrary rank. To bound the top-k sum $F_k(q) = \sum_{j=1}^{k} q_{(j)}$, choose threshold $t = 1/(4k)$ and define the separable upper potential $G_k(q) = \frac{1}{4} + \sum_{i=1}^{N}(q_i - t)_+$, along with the residual discrepancy $D_k(q) = G_k(q) - F_k(q)$. By the Ogryczak–Tamir threshold identity for top-k sums (Ogryczak & Tamir, 2003), $D_k(q) \geq 0$ on Δ_N. Counting the coordinates exceeding t also gives $D_k(q) \leq 1/4$ on the entire simplex (lemma E.1). Taking expectations separates F_k into a report-determined functional and an unresolved residual in $[0, 1/4]$. Combining this with elementary bounds ($F_k \leq kF_1$ and uniform residual sharing) proves that for any report s with $J_k(s) = \sum_{i=1}^{N} \int (x - t)_+ \, \mathrm{d}\xi_i$:

$$R_k \in \big[J_k(s),\ J_k(s) + 1/4\big] \ \cap \ \left[r + \frac{k-1}{N-1}(1 - r),\ \ \min(kr, 1)\right]. \tag{12}$$

For a 10-class posterior report with top-1 accuracy $r = 0.98$, the accuracy-only part of Equation (12) gives $R_2 \in [221/225, 1]$, with width $4/225 \approx 1.78$ percentage points. This narrow interval still straddles a 99% coverage target; certifying that target requires the report-specific lower endpoint $L_2(s) \geq 0.99$.

Exact sharp certificates at ranks two and three. To obtain the tight constants $E_2 = 1/5$ and $E_3 = 17/76$, we construct tighter dual potentials:

$$F_k(q) = \beta_k F_1(q) + \sum_{i=1}^{N} \phi_k(q_i) + h_k(q), \qquad 0 \leq h_k(q) \leq c_k, \quad k \in \{2, 3\}. \tag{13}$$

Here, β_k and the 1-D potential ϕ_k are determined by the report; all ambiguity resides in $\mathbb{E}[h_k(Q)] \in [0, c_k]$. Contact conditions on extremal simplex vertices uniquely pin down the optimal multipliers: $(\beta_2, c_2) = (2/5, 1/5)$ and $(\beta_3, c_3) = (6/19, 17/76)$. Rigorous verification that $0 \leq h_k \leq c_k$ holds across all rows of Δ_N is detailed in Sections C and E.4.

Constructive matching lower bounds. For even $N \geq 4$, averaging two laws on five base rows over N cyclic shifts yields at most $5N$ support points per law, matches all labeled marginals, and preserves the common R_1 and gap

$$R_2^{\mathsf{H}} - R_2^{\mathsf{L}} = \frac{1}{5} - \frac{4}{5N}, \tag{14}$$

approaching $1/5$ as $N \to \infty$. A corresponding rational family matches the $17/76$ certificate for $k = 3$ and lower-bounds general k (Sections D and E.3). Furthermore, the anchor-lift extension demonstrates that even retaining every lower-rank Bayes accuracy $(R_1, \ldots, R_{k-1})$ still fails to identify R_k (proposition F.3).

4.1 From Coverage Intervals to Downstream Budget Decisions

While global constants provide universal worst-case limits, a practitioner evaluating a specific model report needs sharp, instance-specific bounds.

Proposition 4.2 (Attainable report intervals and dual certificates)**.** *For any feasible posterior report $s = (N, \boldsymbol{\xi}, r)$ and rank $1 \leq k \leq N$, the set of attainable top-k coverages is a closed interval $[L_k(s), U_k(s)]$. Let $G(q) = a + \beta F_1(q) + \sum_{i=1}^{N} f_i(q_i)$ for $a, \beta \in \mathbb{R}$ and $f_i \in C([0, 1])$, and define the linear functional $J_s(G) = a + \beta r + \sum_{i=1}^{N} \int_0^1 f_i \, \mathrm{d}\xi_i$. Then*

$$L_k(s) = \sup_{G \leq F_k \text{ on } \Delta_N} J_s(G), \qquad U_k(s) = \inf_{G \geq F_k \text{ on } \Delta_N} J_s(G). \tag{15}$$

Both primal endpoints $L_k(s)$ and $U_k(s)$ are attained by compatible measures in $\mathcal{C}(s)$. The endpoint curves are nondecreasing in k, satisfying $L_1 = U_1 = r$ and $L_N = U_N = 1$.

Proof sketch. The compatibility set $\mathcal{C}(s)$ is a convex, weak-$*$ compact subset of $\mathcal{P}(\Delta_N)$ defined by continuous linear constraints. Since F_k is continuous and affine on $\mathcal{P}(\Delta_N)$, its image is a compact interval. Strong duality follows by specializing constrained-marginal transport duality (Zaev, 2015, Theorem 2.1); see Section G.1. □

Corollary 4.3 (Extremal shortlist candidate budgets)**.** *For any target coverage $\tau \in (0, 1]$, define the optimistic and certified minimum shortlist sizes:*

$$b_{\text{possible}}(s, \tau) = \min\{k : U_k(s) \geq \tau\}, \qquad b_{\text{certified}}(s, \tau) = \min\{k : L_k(s) \geq \tau\}. \tag{16}$$

These bounds represent the exact minimum and maximum of the true budget $b_\tau(\mu) = \min\{k : R_k(\mu) \geq \tau\}$ across all compatible laws $\mu \in \mathcal{C}(s)$. In particular, the shortlist budget is uniquely identified if and only if no coverage interval straddles the target τ:

$$b_{\text{certified}}(s, \tau) - b_{\text{possible}}(s, \tau) = \sum_{k=1}^{N-1} \mathbf{1}\{L_k(s) < \tau \leq U_k(s)\}. \tag{17}$$

corollary 4.3 bridges statistical ambiguity and operational deployment: even when coverage width $W_k(s) > 0$ is positive, the downstream shortlist decision remains unambiguous whenever the target τ falls outside the interval $(L_k(s), U_k(s)]$. Integrating over all targets reveals that total decision ambiguity across the unit interval equals the sum of coverage widths: $\int_0^1 [b_{\text{certified}}(\tau) - b_{\text{possible}}(\tau)]\, \mathrm{d}\tau = \sum_{k=1}^{N-1}(U_k - L_k)$.

5 What Finite Reports Retain in Practice

We now construct reports from trained classifiers and ask what shortlist budget a recipient can infer without access to the evaluation batch. These scores need not be calibrated, so we certify actual label-hit coverage over laws and fixed-size batches compatible with the report. The resulting budgets describe what the report supports, without asserting coverage on future unseen data.

Formulating exact empirical coverage programs. Consider an arbitrary score/label report $\rho = (N, (\gamma_i)_{i=1}^N, a_1)$ as defined in Equation (2), supported on a finite domain $\mathcal{D}_\rho \subset \Delta_N \times [N]$. Rather than relying on calibration assumptions, we evaluate the *actual label-hit probability* under candidate joint assignments π:

$$\begin{aligned} A_k(\pi) &= \sum_{(s,y)\in\mathcal{D}_\rho} \pi_{s,y} \mathbf{1}\{y \in \mathrm{Top}_k(s)\}, \\ \ell_k(\rho) &= \min_{\pi\in\mathcal{P}(\rho)} A_k(\pi), \qquad u_k(\rho) = \max_{\pi\in\mathcal{P}(\rho)} A_k(\pi), \\ \mathcal{P}(\rho) &= \Big\{\pi \geq 0 : \sum_{(s,y)} \pi_{s,y} = 1,\ \mathrm{Law}_\pi\big(S_i, \mathbf{1}\{X = i\}\big) = \gamma_i\ \forall i,\ A_1(\pi) = a_1\Big\}. \end{aligned} \tag{18}$$

When weights $\pi_{s,y} \in [0, 1]$ are continuous, $\ell_k(\rho)$ and $u_k(\rho)$ are solved by standard linear programs (the *weighted-law* setting). For an evaluation batch of fixed size B, imposing the integrality constraint $B\pi_{s,y} \in \mathbb{Z}_{\geq 0}$ yields the exact attainable extrema over *integer batches*. The budget characterization in corollary 4.3 applies directly to either envelope. Exact empirical scores can recover row joins through normalization (proposition G.1), whereas coarse scores can create ambiguity, beyond making computation tractable.

Benchmark experimental protocol. Our empirical evaluation constructs 150 distinct reports spanning five standard tabular and vision datasets (Iris, Wine, Digits, Dry Bean, and Landsat), three classifier families (Random Forest, Extra Trees, and 5-Nearest Neighbors), and ten stratified 70/30 train/test splits. Model predictions use native five-vote or five-neighbor scores without artificial rounding, training-only preprocessors, and ascending-index tie-breaking (Section H). The optimization solvers receive only the anonymized per-class reports ρ; true instance-level test batches are held out solely for auditing. While initial benchmark targets (90%, 95%, 99%) were pre-specified, the full target sweep, summary enrichment, and tie sensitivity are post-hoc analyses.

Per-class reports compress uncertainty, but decision value is target-sensitive. Compared to a naive top-1 baseline that only asserts $A_k \in [A_1, 1]$, providing per-class score/label distributions narrows top-2 coverage width by 51% to 85% on average across non-degenerate dataset/model cells (Table 12). On simple benchmarks (Iris and Wine), coverage is uniquely identified ($W_k = 0$). Across

Table 2: **Shortlist decision ambiguity across 150 benchmark evaluation reports.** A report is decision-ambiguous at target coverage τ when its certified minimum budget $b_{\text{certified}}$ strictly exceeds the optimistic minimum b_{possible}. While all reports identify a unique candidate budget at 90% target, higher targets exhibit substantial ambiguity, requiring up to 3 or 5 additional candidate items.

Target	Top-1 below target	Ambiguous among these	All reports	Max gap
90%	24	0/24	0/150	0
95%	110	29/110	29/150	3
99%	143	85/143	85/150	5

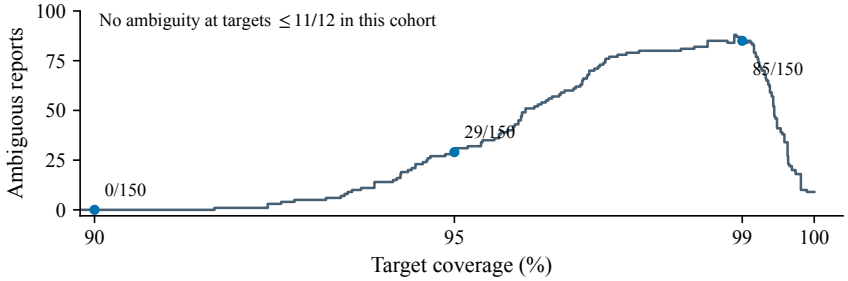


Figure 2: **Decision ambiguity emerges at higher coverage targets.** Exact target sweep of budget gaps ($b_{\text{certified}} - b_{\text{possible}}$) across all 150 benchmark reports. A report becomes ambiguous when at least one coverage interval straddles the target τ (Equation (17)). While no report is ambiguous below $\tau \leq 11/12 \approx 91.7\%$, ambiguity rises steeply toward high targets, with 29 reports ambiguous at 95% and 85 at 99%. Full interval details appear in Section J.

the full cohort, 499 of 570 nontrivial report/rank pairs exhibit positive ambiguity width, including 409 of 420 pairs with $k > 2$.

Crucially, unresolved coverage does not necessarily imply an ambiguous operational decision. As shown in Table 2:

- At a 90% **target**, every single report uniquely identifies the required shortlist size ($b_{\text{certified}} = b_{\text{possible}}$), including all 24 reports where top-1 accuracy is below 90% and multiple candidates are strictly required.
- At a 95% **target**, 29 of the 150 reports leave the candidate budget unresolved, swinging shortlist requirements by up to 3 items.
- At a 99% **target**, 85 reports become decision-ambiguous, with budget discrepancies reaching up to 5 candidates.

Locating missing information via the full target sweep. Sweeping all possible targets $\tau \in [0, 1]$ reveals that 90 out of 150 reports experience decision ambiguity at some threshold. Remarkably, *no report* in this cohort is ambiguous for any target at or below $\tau \leq 11/12 \approx 91.67\%$. As illustrated in Figure 2, decision ambiguity concentrates predominantly in the high-coverage regime ($\tau > 92\%$), where candidate shortlists must capture rare tail errors. In our illustrative Digits/RF report, the top-2 upper bound exceeds 95% ($U_2 = 527/540 \approx 97.6\%$) while the lower bound for four candidates falls short ($L_4 = 128/135 \approx 94.8\%$), leaving the certified budget at five candidates ($L_5 = 527/540$). The ground-truth test batch actually requires three candidates (Figure 1). All 1,740 integer coverage extrema and 900 budget bounds are mathematically verified in Section I.

What additional summary statistics recover. Which additional summaries help the evaluator communicate the shortlist requirement? In Table 3, we examine progressive enrichments for our illustrative report. Adding classwise score laws shrinks the 95% weighted-law budget range from 2–10 down to 2–5. Adding the $N \times N$ argmax confusion matrix further tightens the budget to 2–4. Providing the histogram of the maximum score reduces the total integrated ambiguity without changing the discrete budget bracket. Finally, the cumulative sums of the 1-D *true-label rank*

Table 3: **Progressive recovery of top-k decisions via report enrichment.** Evaluating the illustrative batch under sequentially enriched summary statistics. Each row adds a statistic to all preceding rows. Budget pairs $[b_{\text{possible}}, b_{\text{certified}}]$ represent the attainable range over weighted laws. The integrated gap is the sum of coverage interval widths $\sum_k (u_k - \ell_k)$. For this batch, the true-label rank histogram is sufficient to identify the decision ($b = 3$ at 95%) and is the only tested enrichment level with zero ambiguity. Details: Section J.2.

Information retained	95% budgets	99% budgets	Integrated gap
Top-1 and class frequencies	$(2, 10)$	$(2, 10)$	$2/3$
+ Classwise score/label laws	$(2, 5)$	$(8, 10)$	$5/27$
+ Argmax confusion matrix	$(2, 4)$	$(8, 10)$	$41/270$
+ Maximum-score histogram	$(2, 4)$	$(8, 10)$	$13/90$
+ True-label rank histogram	$(3, 3)$	$(9, 9)$	0

histogram determine every top-k coverage and hence every shortlist budget. For this batch, they recover budgets of 3 at 95% and 9 at 99% without individual evaluation records. This gives a compact reporting option when shortlist sizing is an anticipated use. All 100 enrichment optimization endpoints are certified by exact primal/dual pairs (Section J.2).

Sensitivity to tie-breaking conventions. On the illustrative Digits/RF batch, discrete vote scores induce score ties: zero-score ties cross the top-2 threshold in 41.67% of rows and cross top-5 in 98.89% (Table 15). To assess sensitivity to the tie convention, we recompute reports for this batch under three fixed priority orders: ascending, descending, and cyclic class priority. Across these rules, the 95% budget range spans 2–5 (ascending), 2–4 (descending), and 2–4 (cyclic). Decision ambiguity persists for all three tested priorities on this batch; robustness to other priorities or across the cohort remains unestablished (Section J.4).

Neural networks and score precision. Finally, we study a modern neural model: a pre-trained RepVGG-A0 evaluated on all 10,000 test images of CIFAR-10 (Ding et al., 2021; Krizhevsky, 2009). To study score resolution within our exhaustive implementation, we apply largest-remainder rounding to grid multiples of $1/5$. This maintains normalization but alters ranking dynamics: true top-2 accuracy shifts from 98.35% to 96.43%, and the empirical 99% budget increases from 3 to 7 candidates. Under this rounded report, the certified 99% budget spans six to nine candidates. Admissible states grow from 20,020 at grid $M = 5$ to over 90 million at $M = 20$ (Section K), exceeding our implementation's resource limits. Efficient exact certification for dense, unrounded scores is not established here.

6 Conclusion

Choosing a shortlist for an expensive second stage requires knowing how often the correct label survives the first. Top-1 accuracy and complete per-class diagnostics can leave this coverage undetermined when they do not determine cross-class score associations. For posterior probabilities, we establish sharp dimension-uniform coverage gaps of $1/5$ at rank two and $17/76$ at rank three, converging to $1/4$ as shortlist size and class count grow.

Attainable coverage intervals determine when this missing information changes the candidate budget. In our 150 finite-resolution benchmark reports, all budgets are identified at 90% coverage, while 29 remain ambiguous at 95% and 85 at 99%. One report admits batches requiring either two or five candidates at 95%: the same evaluation summary leaves different amounts of downstream work necessary to meet the target. Argmax confusion matrices and maximum-score histograms reduce ambiguity in this example, but do not resolve every budget.

For evaluators, the practical lesson is to report information suited to the downstream decision. A true-label rank histogram, or equivalently the full top-k coverage curve, determines every fixed-size shortlist budget without releasing individual records. For users, report-compatible budget bounds separate decisions supported by the available evidence from those requiring additional rank information. Extending exact certification to dense, unrounded neural scores and adding finite-sample population guarantees remain open directions.

AI USE STATEMENT

The central subject of this work—information lost through top-1 and marginal evaluation reporting—was first identified with the assistance of a generative AI tool during the analysis of a negative experimental result in a separate project. The question was recognized interesting and investigated as an independent work.

Generative AI tools were then used, under the guidance of the authors, to assist with literature review, problem formalization, proof development and verification, and the implementation of reproducible solver pipelines and exact-arithmetic audit scripts. AI assistance was also used for language refinement and readability editing.

All theoretical results, proofs, empirical data, code implementations, citations, and manuscript revisions were independently verified, audited, and approved by the authors. The authors assume full responsibility for the content, accuracy, and claims of this work.

REPRODUCIBILITY STATEMENT

All theoretical results, sharp constants, and matching extremal constructions are rigorously established in the mathematical appendix (Sections C to E, G.1 and G.2). The accompanying code and computational supplement provide full replication assets:

- **Exact rational-arithmetic checkers:** Independent scripts verify all closed-form bounds, certificates, and potential bounds without floating-point error (`check_calibration.py`, `run_core_checks.py`, `check_topk_theory.py`).
- **Empirical benchmark suites:** All 150 benchmark reports, raw dataset splits, exact integer and weighted-law LP solutions, and 900 budget certificates are documented and cross-verified via SHA-256 hashes (Sections H and I).
- **Enrichment and sensitivity audits:** Complete protocols and reproducible runners for target sweeps, report enrichment (100 endpoints), and tie-rule sensitivity (60 endpoints) are provided in Sections J, J.2 and J.4.
- **Neural audit protocol:** Checkpoint hashes, pre-inference specifications, and certificate verification logs for RepVGG-A0 on CIFAR-10 are detailed in Section K.

Run `verify_package.py` from the extracted supplement before other checks. Complete verification commands, environment specifications, and replication instructions are in its README and Section L. Optional profile, codebook, and realization extensions retain full proofs in Sections M, O and P.

REFERENCES

Stefan Aeberhard and M. Forina. Wine, 1992. URL https://doi.org/10.24432/C5PC7J. Dataset. Bundled scikit-learn 1.5.2 copy used.

Alexander A. Alemi, Ian Fischer, Joshua V. Dillon, and Kevin Murphy. Deep variational information bottleneck. In *International Conference on Learning Representations*, 2017. URL https://openreview.net/forum?id=HyxQzBceg.

E. Alpaydin and C. Kaynak. Optical recognition of handwritten digits, 1998. URL https://doi.org/10.24432/C50P49. Dataset.

Mário S. Alvim, Kostas Chatzikokolakis, Catuscia Palamidessi, and Geoffrey Smith. Measuring information leakage using generalized gain functions. In *2012 IEEE 25th Computer Security Foundations Symposium*, pp. 265–279. IEEE, 2012. doi: 10.1109/CSF.2012.26.

Anastasios Nikolas Angelopoulos, Stephen Bates, Adam Fisch, Lihua Lei, and Tal Schuster. Conformal risk control. In *International Conference on Learning Representations*. OpenReview.net, 2024. URL https://openreview.net/forum?id=33XGfHLtZg.

Omer Antverg and Yonatan Belinkov. On the pitfalls of analyzing individual neurons in language models. In *International Conference on Learning Representations*. OpenReview.net, 2022. URL https://openreview.net/forum?id=8uz0EWPQIMu.

Ehsan Asadi Kangarshahi and Albert Guillén i Fàbregas. Minimum probability of error of list M-ary hypothesis testing. *Information and Inference: A Journal of the IMA*, 12(3):1248–1266, 2023. doi: 10.1093/imaiai/iaad001.

David Blackwell. Equivalent comparisons of experiments. *The Annals of Mathematical Statistics*, 24 (2):265–272, 1953. doi: 10.1214/aoms/1177729032.

Yaofo Chen. PyTorch CIFAR Models. https://github.com/chenyaofo/pytorch-cifar-models, 2026. Source revision 786c16252c0fc58ee9adac063f8337cc4a7a497a; accessed September 13, 2026.

Muthu Chidambaram and Rong Ge. Reassessing how to compare and improve the calibration of machine learning models. In Y. Yue, A. Garg, N. Peng, F. Sha, and R. Yu (eds.), *International Conference on Learning Representations*, volume 2025, pp. 61542–61570, 2025. URL https://proceedings.iclr.cc/paper_files/paper/2025/file/9a30247fcca95299c2fd48d4667282de-Paper-Conference.pdf.

Maxime Darrin, Philippe Formont, Ismail Ben Ayed, Jackie Chi Kit Cheung, and Pablo Piantanida. When is an embedding model more promising than another? In *Advances in Neural Information Processing Systems*, volume 37, pp. 68330–68379, 2024. doi: 10.52202/079017-2183.

Xiaohan Ding, Xiangyu Zhang, Ningning Ma, Jungong Han, Guiguang Ding, and Jian Sun. RepVGG: Making VGG-style convnets great again. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 13733–13742, 2021. URL https://arxiv.org/abs/2101.03697.

Ally Du, Dung Daniel Ngo, and Steven Wu. Reconciling model multiplicity for downstream decision making. In Y. Yue, A. Garg, N. Peng, F. Sha, and R. Yu (eds.), *International Conference on Learning Representations*, volume 2025, pp. 59133–59164, 2025. URL https://proceedings.iclr.cc/paper_files/paper/2025/file/948552777302d3abf92415b1d7e9de70-Paper-Conference.pdf.

Yann Dubois, Douwe Kiela, David J. Schwab, and Ramakrishna Vedantam. Learning optimal representations with the decodable information bottleneck. In *Advances in Neural Information Processing Systems*, volume 33, 2020.

John C. Duchi, Khashayar Khosravi, and Feng Ruan. Multiclass classification, information, divergence, and surrogate risk. *The Annals of Statistics*, 46(6B):3246–3275, 2018. doi: 10.1214/17-AOS1657.

R. A. Fisher. Iris, 1936. URL https://doi.org/10.24432/C56C76. Dataset. Bundled scikit-learn 1.5.2 copy used.

Futoshi Futami and Atsushi Nitanda. Smooth calibration error: Uniform convergence and functional gradient analysis. In C. Vondrick, B. Hariharan, C. Raffel, L. Pinto, D. Yang, and A. Faust (eds.), *International Conference on Learning Representations*, volume 2026, pp. 117–167, 2026. URL https://proceedings.iclr.cc/paper_files/paper/2026/file/00a5213b3479c9733eb3743459badc6b-Paper-Conference.pdf.

Chuan Guo, Geoff Pleiss, Yu Sun, and Kilian Q. Weinberger. On calibration of modern neural networks. In *Proceedings of the 34th International Conference on Machine Learning*, volume 70 of *Proceedings of Machine Learning Research*, pp. 1321–1330. PMLR, 2017. URL https://proceedings.mlr.press/v70/guo17a.html.

Felix Hausdorff. Momentprobleme für ein endliches Intervall. *Mathematische Zeitschrift*, 16(1): 220–248, 1923. doi: 10.1007/BF01175684.

Paul F. Jaeger, Carsten T. Lüth, Lukas Klein, and Till J. Bungert. A call to reflect on evaluation practices for failure detection in image classification. In *International Conference on Learning Representations*. OpenReview.net, 2023. URL https://openreview.net/forum?id=YnkGMIh0gvX.

Piotr Jaworski and Tomasz Rychlik. On distributions of order statistics for absolutely continuous copulas with applications to reliability. *Kybernetika*, 44(6):757–776, 2008. URL https://dml.cz/handle/10338.dmlcz/135889.

Hans G. Kellerer. Duality theorems for marginal problems. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 67(4):399–432, 1984. doi: 10.1007/BF00532047.

Young-Bum Kim, Dongchan Kim, Joo-Kyung Kim, and Ruhi Sarikaya. A scalable neural shortlisting-reranking approach for large-scale domain classification in natural language understanding. In Srinivas Bangalore, Jennifer Chu-Carroll, and Yunyao Li (eds.), *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 3 (Industry Papers)*, pp. 16–24, New Orleans - Louisiana, June 2018. Association for Computational Linguistics. doi: 10.18653/v1/N18-3003. URL https://aclanthology.org/N18-3003/.

Takaaki Koike, Liyuan Lin, and Ruodu Wang. Joint mixability and notions of negative dependence. *Mathematics of Operations Research*, 49(4):2786–2802, 2024. doi: 10.1287/moor.2022.0121.

Alex Krizhevsky. Learning multiple layers of features from tiny images. Technical report, University of Toronto, 2009. URL https://www.cs.toronto.edu/~kriz/learning-features-2009-TR.pdf.

Gowtham R. Kurri, Oliver Kosut, and Lalitha Sankar. Evaluating multiple guesses by an adversary via a tunable loss function. In *2021 IEEE International Symposium on Information Theory*, pp. 2002–2007. IEEE, 2021. doi: 10.1109/ISIT45174.2021.9517733.

Minghan Li, Xinyu Zhang, Ji Xin, Hongyang Zhang, and Jimmy Lin. Certified error control of candidate set pruning for two-stage relevance ranking. In Yoav Goldberg, Zornitsa Kozareva, and Yue Zhang (eds.), *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, pp. 333–345, Abu Dhabi, United Arab Emirates, December 2022. Association for Computational Linguistics. doi: 10.18653/v1/2022.emnlp-main.23. URL https://aclanthology.org/2022.emnlp-main.23/.

Charles Lovering, Rohan Jha, Tal Linzen, and Ellie Pavlick. Predicting inductive biases of pre-trained models. In *International Conference on Learning Representations*. OpenReview.net, 2021. URL https://openreview.net/forum?id=mNtmhaDkAr.

Annabelle McIver, Carroll Morgan, Geoffrey Smith, Barbara Espinoza, and Larissa Meinicke. Abstract channels and their robust information-leakage ordering. In Martín Abadi and Steve Kremer (eds.), *Principles of Security and Trust*, volume 8414 of *Lecture Notes in Computer Science*, pp. 83–102. Springer, 2014. doi: 10.1007/978-3-642-54792-8_5.

Włodzimierz Ogryczak and Arie Tamir. Minimizing the sum of the k largest functions in linear time. *Information Processing Letters*, 85(3):117–122, 2003. doi: 10.1016/S0020-0190(02)00370-8.

Giovanni Puccetti, Bin Wang, and Ruodu Wang. Advances in complete mixability. *Journal of Applied Probability*, 49(2):430–440, 2012. doi: 10.1239/jap/1339878796.

Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, Gretchen Krueger, and Ilya Sutskever. Learning transferable visual models from natural language supervision. In Marina Meila and Tong Zhang (eds.), *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pp. 8748–8763. PMLR, 2021. URL https://proceedings.mlr.press/v139/radford21a.html.

Tomasz Rychlik. Bounds for expectation of L-estimates for dependent samples. *Statistics*, 24(1):1–7, 1993. doi: 10.1080/02331888308802385.

Tomasz Rychlik. Distributions and expectations of order statistics for possibly dependent random variables. *Journal of Multivariate Analysis*, 48(1):31–42, 1994. doi: 10.1016/0047-259X(94)80003-E.

Yuta Sakai. Generalizations of Fano's inequality for conditional information measures via majorization theory. *Entropy*, 22(3):288, 2020. doi: 10.3390/e22030288.

Igal Sason and Sergio Verdú. Arimoto–Rényi conditional entropy and Bayesian M-ary hypothesis testing. *IEEE Transactions on Information Theory*, 64(1):4–25, 2018. doi: 10.1109/TIT.2017.2757496.

Scikit-learn developers. load_digits: Dataset documentation, version 1.5.2. Scikit-learn documentation, 2024. URL https://scikit-learn.org/1.5/modules/generated/sklearn.datasets.load_digits.html. Accessed September 8, 2026.

Scikit-learn developers. ExtraTreesClassifier: Api documentation, version 1.8.0. Scikit-learn documentation, 2025a. URL https://scikit-learn.org/1.8/modules/generated/sklearn.ensemble.ExtraTreesClassifier.html. Accessed September 11, 2026.

Scikit-learn developers. RandomForestClassifier: Api documentation, version 1.8.0. Scikit-learn documentation, 2025b. URL https://scikit-learn.org/1.8/modules/generated/sklearn.ensemble.RandomForestClassifier.html. Accessed September 11, 2026.

Scikit-learn developers. KNeighborsClassifier: Api documentation, version 1.8.0. Scikit-learn documentation, 2025c. URL https://scikit-learn.org/1.8/modules/generated/sklearn.neighbors.KNeighborsClassifier.html. Accessed September 11, 2026.

Ashwin Srinivasan. Statlog (Landsat Satellite), 1993. URL https://doi.org/10.24432/C55887. Dataset. Accessed September 11, 2026.

Jingwu Tang, Jiayun Wu, Steven Wu, and Jiahao Zhang. Dimension-free decision calibration for nonlinear loss functions. In C. Vondrick, B. Hariharan, C. Raffel, L. Pinto, D. Yang, and A. Faust (eds.), *International Conference on Learning Representations*, volume 2026, pp. 86293–86321, 2026. URL https://proceedings.iclr.cc/paper_files/paper/2026/file/8ad060589ed7e18aea4ee1cc8a486c89-Paper-Conference.pdf.

TensorFlow. TensorFlow Model Analysis Architecture. TFX documentation, 2022. URL https://www.tensorflow.org/tfx/model_analysis/architecture. Updated May 20, 2022. Accessed September 5, 2026.

TensorFlow. TensorFlow Model Analysis Metrics and Plots. TFX documentation, 2024. URL https://www.tensorflow.org/tfx/model_analysis/metrics. Updated August 20, 2024. Accessed September 5, 2026.

Marco Tomamichel, Mario Berta, and Masahito Hayashi. Relating different quantum generalizations of the conditional Rényi entropy. *Journal of Mathematical Physics*, 55(8):082206, 2014. doi: 10.1063/1.4892761.

Michael Tschannen, Josip Djolonga, Paul K. Rubenstein, Sylvain Gelly, and Mario Lucic. On mutual information maximization for representation learning. In *International Conference on Learning Representations*. OpenReview.net, 2020. URL https://openreview.net/forum?id=rkxoh24FPH.

UCI Machine Learning Repository. Dry bean, 2020. URL https://doi.org/10.24432/C50S4B. Dataset. Accessed September 11, 2026.

Neria Uzan and Nir Weinberger. A representation-learning game for classes of prediction tasks. In *International Conference on Learning Representations*. OpenReview.net, 2024. URL https://openreview.net/forum?id=Uw8xvFqVAE.

WhyLabs. whylogs: Profiles. whylogs documentation, version 1.6.4, n.d. URL https://docs.whylogs.com/en/latest/index.html. Accessed September 5, 2026.

Yilun Xu, Shengjia Zhao, Jiaming Song, Russell Stewart, and Stefano Ermon. A theory of usable information under computational constraints. In *International Conference on Learning Representations*, 2020. URL https://openreview.net/forum?id=r1eBeyHFDH.

Forest Yang and Sanmi Koyejo. On the consistency of top-k surrogate losses. In *Proceedings of the 37th International Conference on Machine Learning*, volume 119 of *Proceedings of Machine Learning Research*, pp. 10727–10735. PMLR, 2020. URL https://proceedings.mlr.press/v119/yang20f.html.

Danila Zaev. On the Monge–Kantorovich problem with additional linear constraints. *Mathematical Notes*, 98:725–741, 2015. doi: 10.1134/S0001434615110036.

Shengjia Zhao, Abhishek Sinha, Yutong He, Aidan Perreault, Jiaming Song, and Stefano Ermon. Comparing distributions by measuring differences that affect decision making. In *International Conference on Learning Representations*. OpenReview.net, 2022. URL https://openreview.net/forum?id=KB5onONJIAU.

APPENDIX NAVIGATION

The essential proof path is the main theorem, residual bounds, matching families, and general-rank construction (Block I). Blocks II and III give report-specific optimization and empirical certificates. Block IV contains optional profile, equal-mass codebook, and common-prior extensions; none is needed for the main information-limit or budget claims. The table points to the argument and saved checker for each main claim. Checkers verify the stated finite certificates; the continuum claims also require the proofs.

Main claim	Proof or construction	Checker in `code/`
Nonidentification at every N, k	Section E.2	`check_topk_theory.py`
Sharp top-two constant	Potential: Equation (26); residual: Section C; matching family: Section D	`run_core_checks.py`
Sharp top-three constant	Potential: Equation (64); analytic reduction and vertices: Section E.4; matching family: Section E.3	`check_topk_theory.py`; `check_top3_vertices.py`
Quarter limit	Residual proof: Section E.1; finite family: Section E.3	`check_topk_theory.py`
Ten-class top-two bracket	Sections F.2 and F.3	`check_finite_grid_lp.py`; `check_fixed_n_upper.py`
Exact report intervals	Sections G.1 and G.2	`check_report_endpoints.py`
Classical coverage and budgets	Sections H and I; budget proof: corollary 4.3	`shortlist_budget_study.py`
All-target budget identity	Main-text Equation (17); exact cells: Section J	`target_sweep_20260918/analyze.py`
Frozen-score enrichment	Section J.2; all 100 endpoints	`report_enrichment_20260918/study.py` (`check`)
Fixed tie-rule sensitivity	Section J.4; 60 weighted-law endpoints	`tie_sensitivity_20260918/study.py` (`check`)
Rounded neural intervals and budgets	Section K	In `neural_resolution_20260913/`: `analyze_study.py`; `check_neural_budgets.py`

BLOCK I: CORE PROOFS

A POSTERIOR REALIZATION AND CLASSWISE DIAGNOSTICS

List success. Conditioned on $Q = q$, a list $L \subseteq [N]$ with $|L| \le k$ succeeds with probability $\sum_{i \in L} q_i$, maximized by the k largest coordinates. Averaging gives $R_k(\mu) = \int F_k \, \mathrm{d}\mu$.

Bayes realization. For $\mu \in \mathcal{P}(\Delta_N)$, define

$$\Pr\{X = i, Q \in A\} = \int_A q_i \, \mu(\mathrm{d}q).$$

Summing over i shows that the Q-marginal is μ; the X-marginal is $p_i = \int q_i \, \mu(\mathrm{d}q)$, and disintegration gives $\Pr\{X = i \mid Q = q\} = q_i$. If $p_i = 0$, nonnegativity implies $q_i = 0$ almost surely.

The score assigned to the true label has probability law

$$\lambda_\mu(\mathrm{d}x) := x\nu_\mu(\mathrm{d}x) = \mathrm{Law}(Q_X)(\mathrm{d}x). \tag{19}$$

Given a posterior row, a coordinate of value x is the true label with probability x. This size bias explains the factor x and why λ_μ ignores zero padding. It is the law of the score assigned to the true class, which need not be the largest score.

A.1 CLASSWISE REPORTS AND REUSE ACROSS DECISIONS

A reporting workflow supported by existing tools. A model owner exports top-1 accuracy and classwise diagnostics; a recipient later asks about a two-label shortlist. Obtaining the new metric from the exported report, without another evaluation run, is the motivating question. TFMA supports this separation: its evaluators export aggregates while optionally retaining example-level records (TensorFlow, 2022; 2024). The mergeability of whylogs profiles provides a similar compact-summary pathway (WhyLabs, n.d.). These mechanisms establish the workflow's feasibility, not its prevalence.

What complete classwise diagnostics retain. For a posterior law μ, write $Y_i = \mathbf{1}\{X = i\}$ and $\eta_{\mu,i} = \mathrm{Law}(Q_i, Y_i)$.

Proposition A.1 (Complete one-vs-rest diagnostic laws)**.** *On a common finite label space, the tuple $(\eta_{\mu,i})_{i=1}^N$ and the labeled marginal tuple $\boldsymbol{\xi}_\mu$ determine one another. In particular, matching $\boldsymbol{\xi}$ matches every population diagnostic that is a function of these separate binary laws, including classwise threshold confusion tables, calibration-bin totals, and ROC and precision–recall curves wherever their defining denominators are positive.*

Proof. For Borel $B \subseteq [0, 1]$, conditional expectation gives

$$\mathbb{E}[\mathbf{1}\{Q_i \in B\}Y_i] = \mathbb{E}[\mathbf{1}\{Q_i \in B\}Q_i] = \int_B s \, \xi_{\mu,i}(\mathrm{d}s).$$

Replacing Y_i by $1 - Y_i$ proves the second identity in Equation (4). These identities specify $\eta_{\mu,i}$; conversely, summing its two label components gives $\xi_{\mu,i}$. At threshold τ, true- and false-positive population masses are respectively

$$\mathrm{TP}_i(\tau) = \int_{[\tau,1]} s \, \xi_{\mu,i}(\mathrm{d}s), \qquad \mathrm{FP}_i(\tau) = \int_{[\tau,1]} (1 - s) \, \xi_{\mu,i}(\mathrm{d}s).$$

The prior gives the remaining two confusion-table entries. For a calibration bin B, its population mass is $\xi_{\mu,i}(B)$, and both its label and score totals equal $\int_B s \, \xi_{\mu,i}(\mathrm{d}s)$. Normalizing these quantities gives the stated curves and bin averages when defined. □

The proposition concerns *one-vs-rest* diagnostics examining (Q_i, Y_i) separately. A multiclass argmax confusion matrix, calibration conditioned on the predicted class, and the distribution of the largest score depend on row associations and are not covered. The supplied R_1 is one additional row-based scalar. Bayes multiclass log loss and Brier risk are fixed: $-\sum_i \int s \log s \, \mathrm{d}\xi_{\mu,i}$ and $1 - \sum_i \int s^2 \, \mathrm{d}\xi_{\mu,i}$, respectively ($0 \log 0 = 0$).

		Class 1	Class 2	Class 3	Class 4	R_1	R_2
A	q^1	0	0	1/3	2/3	$\frac{1}{2}$	$\frac{5}{6}$
	q^2	1/6	1/3	1/6	1/3		
B	q^1	1/6	0	1/6	2/3	$\frac{1}{2}$	$\frac{3}{4}$
	q^2	0	1/3	1/3	1/3		

Figure 3: **Same posterior columns, different top-2 coverage.** Each row has probability $1/2$. Column distributions, prior $(1/12, 1/6, 1/4, 1/2)$, and mean maximum $R_1 = 1/2$ agree. Borders mark a top-2 choice. The mean top-two sums are $5/6$ and $3/4$: a gap of $1/12$, below the sharp global limit $1/5$.

Proposition A.2 (Calibrated scores and transfer to arbitrary scores)**.** *Let $S \in \Delta_N$ and define $A_k = \Pr\{X \in \mathrm{Top}_k(S)\}$, using any fixed measurable rule for breaking ties. If $\Pr\{X = i \mid S\} = S_i$ for every i, then $A_k = \mathbb{E}F_k(S) = \mathrm{Acc}^*_k(X \mid S)$, and the diagnostic-law identity holds with S in place of Q. Without this calibration assumption, the class of joint laws of (S, X) still contains, for every $\varepsilon > 0$, finite common-prior pairs matching all $\mathrm{Law}(S_i, \mathbf{1}\{X = i\})$ and A_1, with $|A_2^{\mathsf{A}} - A_2^{\mathsf{B}}| > 1/5 - \varepsilon$. Thus the global minimax expected absolute error based on these reports is at least $1/10$. No matching upper bound for arbitrary scores is asserted.*

Proof. Under the calibration condition, conditioning on $S = s$ makes the success probability of any chosen label list its score sum; a top-k list attains $F_k(s)$, including at ties. The tower property gives $\mathbb{E}(\mathbf{1}\{X = i\} \mid S_i) = S_i$, hence the diagnostic identity. For the second claim, take the finite posterior pairs of theorem B.4, set $S = Q$, and use their Bayes label kernels. Their binary laws agree by proposition A.1, and their actual list accuracies are the corresponding R_k. The common-report two-point triangle inequality gives a loss at least $(1/5 - \varepsilon)/2$; let $\varepsilon \downarrow 0$. □

Classwise calibration does not identify score sums with coverage. Give each of the following score rows probability $1/4$. Every displayed vector is divided by 12; the second column is the true conditional label kernel, not another score vector.

$12S$	$12\Pr\{X = \cdot \mid S\}$
$(0, 0, 4, 8)$	$(0, 0, 5, 7)$
$(2, 4, 2, 4)$	$(2, 4, 3, 3)$
$(2, 0, 2, 8)$	$(2, 0, 1, 9)$
$(0, 4, 4, 4)$	$(0, 4, 3, 5)$

The first two coordinates are unchanged. In the third, perturbations $(+1, +1, -1, -1)/12$ cancel within each equal-score group; in the fourth, $(-1, -1, +1, +1)/12$ do likewise. Thus $\Pr\{X = i \mid S_i\} = S_i$ for every coordinate, although vector calibration fails. With ascending-index ties, the selected top-1 labels are $(4, 2, 4, 2)$ and the top-2 lists are $(\{4, 3\}, \{2, 4\}, \{4, 1\}, \{2, 3\})$. Consequently,

$$A_1 = \frac{7 + 4 + 9 + 4}{48} = \frac{1}{2} = \mathbb{E}F_1,$$
$$A_2 = \frac{12 + 7 + 11 + 7}{48} = \frac{37}{48}, \qquad \mathbb{E}F_2 = \frac{12 + 8 + 10 + 8}{48} = \frac{19}{24}.$$

This directly separates actual coverage from the score-sum objective even when top-1 agrees. The exact check is supplied in `code/check_calibration.py`.

The running pair as a report-reuse example. The two exact laws in Figure 3 share every labeled marginal. By proposition A.1, their entire one-vs-rest diagnostic reports agree, at every score threshold and under any common binning. The originally supplied accuracy is $R_1 = 1/2$ for both, whereas two-label success is $5/6$ versus $3/4$. The near-sharp symmetrized pairs in theorem B.4 have the same diagnostic equivalence and make the top-2 gap approach $1/5$.

Table 4: **Scope of the top-2 guarantees.** Every row includes the original top-1 accuracy. Lower-bound examples are worst-case existence statements, not guarantees for every report.

Scores and retained population report	Conclusion established here
Posteriors or fully vector-calibrated scores; exact labeled marginals	Global diameter $1/5$, minimax error $1/10$; individual intervals may be narrower.
Arbitrary normalized scores; complete separate score/label laws	Examples with gap approaching $1/5$, hence minimax error at least $1/10$; no matching upper bound claimed.
Coarser deterministic classwise reports	The same examples remain indistinguishable; oracle upper bounds need not transfer.

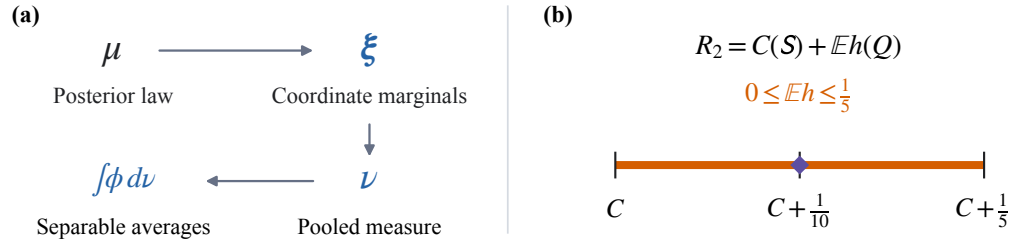


Figure 4: **From coordinate summaries to a decision certificate.** (a) The joint posterior law μ determines the labeled marginals $\boldsymbol{\xi}$, which determine the pooled measure $\nu = \sum_i \xi_i$. The pooled measure fixes every symmetric separable average. (b) With R_1 supplied, the summary fixes $C(\mathsf{S}) = \frac{2}{5}R_1 + \int \phi \, \mathrm{d}\nu$, where ϕ is defined in Section B. The remaining $\mathbb{E}h \in [0, 1/5]$ gives a universal interval of width $1/5$; its diamond-marked midpoint has radius $1/10$. A fixed summary fiber may be narrower.

Proposition A.3 (Passing to a coarser population report)**.** *Let $T(\mu) = g(N, \boldsymbol{\xi}_\mu, R_1(\mu))$ be any deterministic report formed from the stated oracle. For every $\varepsilon > 0$, there exist finite common-uniform-prior laws with the same T and top-2 gap greater than $1/5 - \varepsilon$. Consequently, any possibly randomized estimator using only T has global worst-case expected absolute error at least $1/10$. No corresponding upper bound is asserted for arbitrary g.*

Proof. Apply g to a same-oracle pair from theorem B.4; its reports coincide. An estimator has the same output law under both reports. Integrating the triangle inequality shows that one of the two expected absolute errors is at least half their top-2 gap. Let $\varepsilon \downarrow 0$. □

The proposition covers population histograms with fixed bins and finite classwise threshold panels. Finite-sample reports require a separate observation model: classwise histograms computed from shared examples can retain dependence information through their joint fluctuations.

What to record. When list size k is known, the owner can accumulate the classifier's top-k label-hit rate without storing every row. Complete one-coordinate plots, even when exact, do not generally determine a missing list statistic.

B The Sharp Uniform Limit for Bayes Top-2

Notation for the top-2 certificates. For the report $s = (N, \boldsymbol{\xi}, r)$ of Section 3, write

$$\mathcal{D}^{\mathrm{marg}}_{N,\boldsymbol{\xi},r} = W_2(s). \tag{20}$$

For feasible pooled constraints, define

$$\mathcal{D}_{N,\nu,r} = \sup_{\substack{\nu_{\mu_\mathsf{A}} = \nu_{\mu_\mathsf{B}} = \nu \\ R_1(\mu_\mathsf{A}) = R_1(\mu_\mathsf{B}) = r}} |R_2(\mu_\mathsf{A}) - R_2(\mu_\mathsf{B})|. \tag{21}$$

Pooling relaxes the report, so

$$\mathcal{D}^{\mathrm{marg}}_{N,\boldsymbol{\xi},r} \le \mathcal{D}_{N,\sum_i \xi_i,r}. \tag{22}$$

The pooled family can include different priors. Our lower constructions match every labeled marginal and can share a uniform prior.

The upper bound uses only the pooled measure $\nu_\mu = \sum_i \xi_{\mu,i}$ and the reported top-1 value. We first give a certificate valid in every finite dimension, then establish sharpness with pairs that also match the full labeled marginal tuple.

Theorem B.1 (Universal bound with complete coordinate marginals)**.** *For every finite N and feasible (ν, r),*

$$\boxed{\mathcal{D}_{N,\nu,r} \le \frac{1}{5}.} \tag{23}$$

The posterior laws may be arbitrary Borel laws and need not share a barycenter. Consequently, for every coordinate-feasible $(\boldsymbol{\xi}, r)$,

$$\boxed{\mathcal{D}^{\mathrm{marg}}_{N,\boldsymbol{\xi},r} \le \frac{1}{5}.} \tag{24}$$

The labeled-marginal bound follows from Equation (22). Those pairs share a prior, while the pooled theorem allows a larger family. The constant $1/5$ bounds every feasible report; an individual report can have smaller width. The construction below proves that no smaller constant works uniformly across dimensions and reports.

Designing the separable certificate. Complete labeled marginals cancel coordinate-separable averages. For the pooled summary, the same potential is used in every coordinate.

Proposition B.2 (Dual oscillation principle)**.** *For $c \in \mathbb{R}$ and continuous η with $\eta(0) = 0$, put $u_{c,\eta}(q) = F_2(q) - cF_1(q) - \sum_i \eta(q_i)$. Every feasible slice obeys*

$$\mathcal{D}_{N,\nu,r} \le \sup_{q\in\Delta_N} u_{c,\eta}(q) - \inf_{q\in\Delta_N} u_{c,\eta}(q). \tag{25}$$

Proof. Equal top-1 and intensity cancel the first two retained terms. The remaining difference is between two expectations of $u_{c,\eta}$. □

We seek a coefficient and potential with small residual oscillation. The sharp construction suggests a contact ansatz: normalize the lower residual to zero and impose $\phi(s) = o(s)$ near zero, so a diffuse tail contributes vanishing potential. Contact on paired and separated atoms at $1/2$ and $1/4$ gives

$$d = \frac{c}{2} = \frac{1}{4} - \frac{c}{8}, \qquad c = \frac{2}{5}, \quad d = \frac{1}{5}, \quad \phi(1/4) = \frac{1}{10}, \quad \phi(1/2) = \frac{3}{10}.$$

A quadratic branch near zero and affine interpolation between the contacts are completed by $\phi(s) + \phi(1-s) = 4/5 - 2s/5$ for $s \ge 1/2$, giving the complementary breakpoint $3/4$. These contact conditions suggest a candidate potential. Its validity follows from the uniform residual bound proved below (Section C.1).

Define the continuous atom potential

$$\phi(s) = \begin{cases} \frac{8}{5}s^2, & 0 \le s \le \frac{1}{4}, \\ \frac{4}{5}s - \frac{1}{10}, & \frac{1}{4} \le s \le \frac{1}{2}, \\ \frac{2}{5}s + \frac{1}{10}, & \frac{1}{2} \le s \le \frac{3}{4}, \\ -\frac{4}{5} + \frac{14}{5}s - \frac{8}{5}s^2, & \frac{3}{4} \le s \le 1, \end{cases} \tag{26}$$

and, for $q \in \Delta_N$,

$$h(q) = F_2(q) - \frac{2}{5}F_1(q) - \sum_{i=1}^{N} \phi(q_i). \tag{27}$$

Integration gives the exact decomposition

$$R_2(\mu) = \frac{2}{5}R_1(\mu) + \int \phi \, \mathrm{d}\nu_\mu + \mathbb{E}_\mu h. \tag{28}$$

The report determines the first two terms exactly. Its missing information enters only through the expectation of the nonseparable residual (Figure 4b).

Lemma B.3 (Scalar residual certificate)**.** *For every finite $N \geq 1$ and $q \in \Delta_N$,*

$$0 \leq h(q) \leq \frac{1}{5}. \tag{29}$$

This includes ties, zeros, and $N = 1$, with $q_{(2)} = 0$.

Proof sketch. Put $x = q_{(1)}$, $y = q_{(2)}$, and $g = \phi|_{[0,1/2]}$. The ratio $g(s)/s$ is nondecreasing, so all tail mass can be charged at the runner-up rate. Together with monotonicity of $s - g(s)$, this gives the upper bound and reduces the lower bound to a two-variable piecewise-polynomial check in (x, y). The complete cases, boundaries, ties, and equality sets are in Sections C and O.2. □

Proof of theorem B.1. By Equations (11) and (27),

$$R_2(\mu) = \frac{2}{5} R_1(\mu) + \int \phi \, \mathrm{d}\nu_\mu + \mathbb{E}_\mu h.$$

On one slice, the first two terms cancel between the laws. The remaining difference is between two expectations in $[0, 1/5]$. □

A five-row sharp sequence. For an integer $v \geq 2$, set

$$N_v = 2v, \qquad t = \frac{1}{2v}, \qquad a = \frac{1}{2} - t, \qquad b = \frac{1}{4}.$$

The high side pairs the large atoms; the low side separates them and includes one diffuse row (Equation (49)). Both sides have the same positive inventory $\{a^{[2]}, b^{[8]}, t^{[4v+2]}\}$ and, after padding, the same zero inventory. Equal weighting of the five rows gives a common intensity ν_v. Direct calculation also gives the same top-1 value $r_v = (3/2 - t)/5$, while

$$R_2(\mu_{\mathsf{H},v}) - R_2(\mu_{\mathsf{L},v}) = \frac{1}{5} - \frac{2}{5v}. \tag{30}$$

Figure 5 evaluates these finite constructions: the common top-1 value tends to 0.3, while the top-2 gap grows from 0.15 at $N = 16$ to 0.1875 at $N = 64$. Along this sequence, the remaining deficit is exactly $1/5 - \Delta R_2 = 4/(5N_v)$. The full row and inventory arithmetic is in Section D.

Theorem B.4 (Sharpness with every coordinate marginal known)**.**

$$\sup_{N<\infty} \sup_{(\boldsymbol{\xi},r) \text{ coordinate-feasible}} \mathcal{D}^{\mathrm{marg}}_{N,\boldsymbol{\xi},r} = \sup_{N<\infty} \sup_{(\nu,r) \text{ feasible}} \mathcal{D}_{N,\nu,r} = \frac{1}{5}. \tag{31}$$

For every v, a permutation-symmetrized finite pair has, for every labeled coordinate j, the common marginal law ν_v/N_v, the common top-1 value r_v, and the gap in Equation (30). The pair admits a realization under one common uniform prior.

Proof. Theorem B.1 gives both upper bounds, and Equation (30) gives the reverse inequality for the pooled slices as $v \to \infty$. Uniformly averaging each finite posterior law over coordinate permutations preserves ν and every R_k. Moreover, Equation (59) gives

$$\xi_{\widetilde{\mu}_{\mathsf{H},v},j} = \xi_{\widetilde{\mu}_{\mathsf{L},v},j} = \frac{1}{N_v} \nu_v, \qquad j = 1, \ldots, N_v.$$

Thus the same gaps prove the reverse inequality for the coordinate-marginal slices. Section D also gives the uniform barycenter and the corresponding Bayes kernels. □

Each v-pair shares its dimension, marginals, and top-1 value; these quantities vary along the sequence. The construction proves sharpness of the outer supremum. It does not prescribe a positive gap for a given marginal tuple, profile, or prior.

Finite label spaces. Write $\mathcal{D}_N^\star = \sup_{(\nu,r)\text{ feasible}} \mathcal{D}_{N,\nu,r}$. The five-row construction and certified finite-grid optima give lower bounds. For smaller continuum upper bounds, we choose a piecewise-affine potential and verify its residual at every vertex of the ordered-simplex cells cut by the potential knots. The residual is affine on each cell, so these vertices capture its extrema over all posterior rows, including rows off the grid (proposition F.2). The resulting brackets include

$$\frac{7}{48} \le \mathcal{D}_{10}^\star \le \frac{7}{45}, \qquad \frac{24}{125} \le \mathcal{D}_{100}^\star \le \frac{392}{2013}.$$

Table 5 reports the bounds for common class counts; Sections F.2 and F.3 provide their certificates. These brackets bound the worst case over reports at each dimension. The exact continuum suprema remain open.

Optimal summary-only estimation. The certificate interval and sharp construction together determine the exact global minimax error. For a posterior law μ, let $\mathsf{S}_{\mathrm{marg}}(\mu) = (N, \boldsymbol{\xi}_\mu, R_1(\mu))$. A possibly randomized point rule K assigns a distribution $K(\mathsf{S}_{\mathrm{marg}}, \mathrm{d}\widehat{r})$ on $\mathbb{R}$ to each attainable summary.

Theorem B.5 (Exact coordinate-marginal minimax risk)**.** *Over all posterior laws in all finite label dimensions,*

$$\boxed{\inf_K \sup_{\substack{N<\infty \\ \mu\in\mathcal{P}(\Delta_N)}} \int |\widehat{r} - R_2(\mu)|\, K(\mathsf{S}_{\mathrm{marg}}(\mu),\, \mathrm{d}\widehat{r}) = \frac{1}{10}.} \tag{32}$$

The deterministic midpoint rule

$$T^\star(\mathsf{S}_{\mathrm{marg}}(\mu)) = \frac{2}{5} R_1(\mu) + \int \phi \, \mathrm{d}\nu_\mu + \frac{1}{10} \tag{33}$$

attains the upper bound. The lower bound already holds when the supremum is restricted to finitely supported laws with uniform barycenter.

Proof. The coordinate marginals determine $\nu_\mu = \sum_i \xi_{\mu,i}$, so $T^\star$ uses only $\mathsf{S}_{\mathrm{marg}}$. By Equations (28) and (29), its error is $|1/10 - \mathbb{E}_\mu h| \le 1/10$. Conversely, for every $\varepsilon > 0$, the symmetrized sharp construction gives finite-support uniform-barycenter laws with the same summary and top-2 gap greater than $1/5 - \varepsilon$. A rule has the same output law under both. The triangle inequality forces one expected absolute loss to exceed $1/10 - \varepsilon/2$; let $\varepsilon \downarrow 0$. □

The midpoint of the universal certificate attains the global minimax value. For a supplied report, a narrower interval can yield a smaller conditional minimax error, as characterized in Section 4.1.

The complete positive-order Rényi-profile version is stated separately in corollary M.2. Approximate summaries, longer lists, and equal-output-budget constraints are developed in Sections F, N and Q.

C Complete proof of the scalar residual certificate

C.1 A reconstruction from the contact rows

For a dimension-uniform certificate, a convenient admissible class is

$$\Phi = \{\phi \in C([0,1]) : \phi(0) = 0,\ \sup_{s>0} |\phi(s)|/s < \infty\}.$$

The ratio bound makes $\sum_i \phi(q_i)$ uniformly bounded on every finite simplex, and $\phi(0) = 0$ makes zero padding immaterial. Consider the optimization

$$\inf_{c\in\mathbb{R},\, \phi\in\Phi} \left(\sup_{N, q\in\Delta_N} u_{c,\phi}(q) - \inf_{N, q\in\Delta_N} u_{c,\phi}(q) \right).$$

The potential in Equation (26) attains value $1/5$. A smaller value would contradict proposition B.2 and the matched pairs approaching $1/5$ in theorem B.4. Thus the explicit certificate solves this oscillation problem. The general transport-duality framework remains the one described in Section 2.

To reconstruct the potential, impose $\phi(s) = o(s)$ near zero. For a tail of atoms of size t and bounded total mass, the sum of potential values is bounded in absolute value by a constant multiple of $|\phi(t)|/t$, which vanishes as $t \downarrow 0$. The limiting rows in Equation (49) suggest the contacts below. Write $A = \phi(1/4)$, $B = \phi(1/2)$, and normalize the lower residual to zero and the upper residual to d:

$$\begin{aligned} 1 - c/2 - 2B &= d \quad \text{(two half-mass atoms)}, \\ 1/2 - c/2 - B &= 0 \quad \text{(one half-mass atom and a diffuse tail)}, \\ 1/2 - c/4 - 2A &= d \quad \text{(two quarter-mass atoms and a diffuse tail)}, \\ 1/2 - c/4 - 4A &= 0 \quad \text{(four quarter-mass atoms)}. \end{aligned} \tag{34}$$

The second equation gives $B = (1-c)/2$, so the first gives $d = c/2$. The fourth gives $A = 1/8 - c/16$, so the third gives $d = 1/4 - c/8$. Consequently $c = 2/5$, $d = 1/5$, $A = 1/10$, and $B = 3/10$. These values are forced within this contact ansatz; the equations do not uniquely determine a function between the contacts.

Choose the quadratic $g(s) = 8s^2/5$ on $[0, 1/4]$, which vanishes to first order and takes value A at $1/4$. Interpolate affinely from A to B on $[1/4, 1/2]$, giving $g(s) = 4s/5 - 1/10$. For a binary row $(s, 1-s)$, $s \geq 1/2$, imposing upper contact gives

$$\phi(s) + g(1-s) = 1 - cs - d = 4/5 - 2s/5.$$

Substituting the two branches of g gives the upper branches of Equation (26), with breakpoint $3/4$. The interpolation makes both $g(s)/s$ and $s - g(s)$ nondecreasing. These properties control the tail in the proof below. The contact equations reconstruct a candidate; the residual inequalities establish its validity. This derivation does not provide sharp potentials for arbitrary list functionals.

C.2 Uniform residual bounds

We prove Lemma B.3 without suppressing boundary cases. The formulas in (26) give

$$g(s) = \begin{cases} \dfrac{8}{5}s^2, & 0 \leq s \leq \dfrac{1}{4}, \\ \dfrac{4}{5}s - \dfrac{1}{10}, & \dfrac{1}{4} \leq s \leq \dfrac{1}{2}. \end{cases} \tag{35}$$

Thus $g \geq 0$, and for $s > 0$,

$$\rho(s) := \frac{g(s)}{s} = \begin{cases} \dfrac{8}{5}s, & 0 < s \leq \dfrac{1}{4}, \\ \dfrac{4}{5} - \dfrac{1}{10s}, & \dfrac{1}{4} \leq s \leq \dfrac{1}{2} \end{cases} \tag{36}$$

is continuous at $1/4$. Its derivatives on the two open branches are $8/5$ and $1/(10s^2)$, so it is nondecreasing. Also,

$$\frac{d}{ds}\big(s - g(s)\big) = \begin{cases} 1 - \dfrac{16}{5}s, & 0 < s < \dfrac{1}{4}, \\ \dfrac{1}{5}, & \dfrac{1}{4} < s < \dfrac{1}{2}, \end{cases}$$

so $s - g(s)$ is nondecreasing on $[0, 1/2]$. Direct substitution in the two relevant high branches of ϕ proves

$$\phi(s) + g(1-s) = \frac{4}{5} - \frac{2}{5}s, \qquad \frac{1}{2} \leq s \leq 1. \tag{37}$$

All identities agree at $1/4$, $1/2$, and $3/4$.

Fix $q \in \Delta_N$ and sort its coordinates. Let

$$x := q_{(1)}, \qquad y := q_{(2)}.$$

Our zero-padding convention makes this definition valid also for $N = 1$. We always have

$$0 \leq y \leq \min\{x, 1-x\} \leq \frac{1}{2}. \tag{38}$$

If $y = 0$, normalization forces $q = (1, 0, \ldots, 0)$. Since $\phi(0) = 0$ and $\phi(1) = 2/5$,

$$h(q) = 1 - \frac{2}{5} - \frac{2}{5} = \frac{1}{5}. \tag{39}$$

This includes the entire $N = 1$ case. Henceforth assume $y > 0$.

Every $q_{(i)}$ with $i \geq 2$ lies in $[0, y]$. Monotonicity of ρ, with the zero terms handled by $g(0) = 0$, gives

$$\sum_{i \geq 2} g(q_{(i)}) \leq \rho(y) \sum_{i \geq 2} q_{(i)} = (1 - x)\rho(y). \tag{40}$$

On the other hand, the term $i = 2$ is $g(y)$ and every remaining term is nonnegative, so

$$\sum_{i \geq 2} g(q_{(i)}) \geq g(y). \tag{41}$$

By definition,

$$h(q) = \frac{3}{5}x + y - \phi(x) - \sum_{i \geq 2} g(q_{(i)}). \tag{42}$$

Upper bound. First suppose $x \leq 1/2$. Then $\phi(x) = g(x)$ and $y \leq x$. Using (41) and monotonicity of $s - g(s)$,

$$\begin{aligned} h(q) &\leq \frac{3}{5}x - g(x) + y - g(y) \\ &\leq \frac{3}{5}x - g(x) + x - g(x) \\ &= \begin{cases} \frac{8}{5}x - \frac{16}{5}x^2, & 0 \leq x \leq \frac{1}{4}, \\ \frac{1}{5}, & \frac{1}{4} \leq x \leq \frac{1}{2}. \end{cases} \end{aligned}$$

The quadratic in the first branch equals

$$\frac{1}{5} - \frac{16}{5}\left(x - \frac{1}{4}\right)^2$$

and is therefore at most $1/5$.

Now suppose $x \geq 1/2$. From $y \leq 1 - x$ and the same monotonicity,

$$\begin{aligned} h(q) &\leq \frac{3}{5}x - \phi(x) + y - g(y) \\ &\leq \frac{3}{5}x - \phi(x) + (1 - x) - g(1 - x) \\ &= \frac{1}{5}, \end{aligned}$$

where the last equality is (37). This proves $h(q) \leq 1/5$ on the full domain, including $x = 1/2$.

Lower bound. Equations (40) and (42) show that it is enough to prove

$$L(x, y) := \frac{3}{5}x + y - \phi(x) - (1 - x)\rho(y) \geq 0 \tag{43}$$

for the domain in (38) with $y > 0$.

Case 1: $0 < x \leq 1/4$. Here $0 < y \leq x \leq 1/4$, and

$$L(x, y) = \frac{3}{5}x - \frac{8}{5}x^2 + y\left(\frac{8}{5}x - \frac{3}{5}\right). \tag{44}$$

The coefficient of y is negative, so this expression is minimized at $y = x$, where it is zero.

Case 2: $1/4 \le x \le 1/2$ *and* $0 < y \le 1/4$. In this region,

$$L(x,y) = \frac{1}{10} - \frac{1}{5}x + y\left(\frac{8}{5}x - \frac{3}{5}\right). \tag{45}$$

If $x \le 3/8$, the coefficient of y is nonpositive, so the minimum over $0 < y \le 1/4$ occurs at $y = 1/4$ and equals

$$\frac{1}{5}x - \frac{1}{20} \ge 0.$$

If $x \ge 3/8$, the coefficient is nonnegative and hence

$$L(x,y) \ge \frac{1}{10} - \frac{1}{5}x \ge 0.$$

Case 3: $1/4 \le y \le x \le 1/2$. Multiplying by the positive number $10y$ gives

$$10yL(x,y) = 10y^2 + (6x-7)y + 1 - x. \tag{46}$$

For fixed $y \ge 1/4$, the right-hand side has derivative $6y - 1 > 0$ with respect to x. Since $x \ge y$, it is minimized at $x = y$, where

$$10yL(y,y) = (4y-1)^2 \ge 0.$$

Case 4: $1/2 \le x < 1$. Put $r = 1 - x$. Then $0 < y \le r \le 1/2$, and (37) transforms (43) into

$$L(x,y) = \frac{1}{5} - r + g(r) + y - r\rho(y). \tag{47}$$

There are three subcases.

If $0 < r \le 1/4$, then $y \le 1/4$ and

$$L = \frac{1}{5} - r + \frac{8}{5}r^2 + y\left(1 - \frac{8}{5}r\right).$$

The coefficient of y is positive, while

$$\frac{1}{5} - r + \frac{8}{5}r^2 \ge \frac{1}{20} \qquad (0 \le r \le 1/4).$$

Thus $L > 0$.

If $1/4 \le r \le 1/2$ and $0 < y \le 1/4$, then

$$L = \frac{1}{10} - \frac{1}{5}r + y\left(1 - \frac{8}{5}r\right) \ge 0,$$

because both displayed coefficients are nonnegative on this interval.

Finally, if $1/4 \le y \le r \le 1/2$, then

$$L = \frac{1}{10} - r + y + \frac{r}{10y}. \tag{48}$$

For fixed $r \le 1/2$, its derivative with respect to y satisfies

$$1 - \frac{r}{10y^2} \ge 1 - \frac{1/2}{10(1/4)^2} = \frac{1}{5} > 0.$$

The expression is therefore minimized at $y = 1/4$, where

$$L \ge \frac{7}{20} - \frac{3}{5}r \ge \frac{1}{20}.$$

Cases 1–3 cover $x \le 1/2$, and Case 4 covers $x \ge 1/2$; either adjacent case handles a shared boundary. Hence $L(x,y) \ge 0$ throughout the domain. Combining (40), (42), and (43) gives $h(q) \ge 0$. Together with the upper bound and (39), this proves Lemma B.3. The argument uses only sorted values, so ties require no distinguished maximizing coordinate.

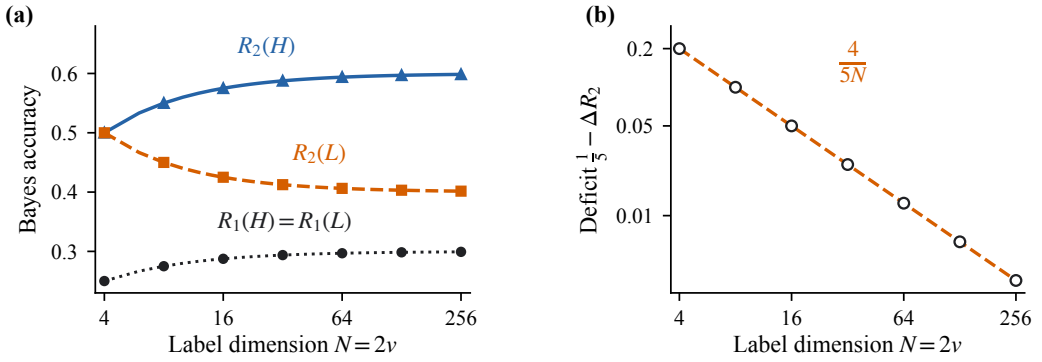


Figure 5: **Finite constructions approach the sharp limit.** (a) Exact top-1 and top-2 accuracies of the five-row laws for even $N = 2v$, $v = 2, \dots, 128$, computed from the posterior rows using rational arithmetic. Averaging over coordinate permutations preserves these accuracies and matches every labeled marginal under a common uniform prior. (b) The deficit from $1/5$ is exactly $4/(5N)$: circles show row-based calculations at selected dimensions and the dashed line shows the formula on logarithmic axes. These are deterministic construction values, not sampled observations or fixed-N optima. Each dimension uses a different matched summary pair.

D THE FIVE-ROW SHARP SEQUENCE AND ITS COMMON-PRIOR REALIZATION

Fix an integer $v \geq 2$ and set $t = 1/(2v)$, $a = 1/2 - t$, $b = 1/4$, and $N_v = 2v$. For clarity, before padding define

$$\begin{aligned} H_1 &= (a, a, t, t), \quad H_2 = H_3 = H_4 = H_5 = (b, b, t^{[v]}), \\ L_1 = L_2 &= (a, t^{[v+1]}), \qquad L_3 = L_4 = (b^{[4]}), \qquad L_5 = (t^{[2v]}). \end{aligned} \tag{49}$$

Row normalization and common dimension. Every displayed row has mass one:

$$2a + 2t = 1, \qquad 2b + vt = 1, \qquad a + (v+1)t = 1, \qquad 4b = 1, \qquad 2vt = 1. \tag{50}$$

The displayed lengths are 4, $v+2$, and $2v$, all at most $N_v = 2v$ for $v \geq 2$. We pad every row on both sides with zeros to this same dimension. In particular, each individual comparison is made inside one common simplex Δ_{N_v}.

Exact zero-inclusive inventory equality. Before padding, the high side has positive-coordinate multiset

$$\{a^{[2]}, b^{[8]}, t^{[4v+2]}\},$$

and the low side has

$$\{a^{[2]}, b^{[8]}, t^{[2(v+1)+2v]}\} = \{a^{[2]}, b^{[8]}, t^{[4v+2]}\}. \tag{51}$$

At $v = 2$, the nominal types collide: $a = b = t = 1/4$. The statement remains literal after coalescing equal values: both sides then contain twenty occurrences of $1/4$.

For every v, each side has $4v + 12$ positive occurrences among $5N_v = 10v$ coordinate slots. Hence each side also has exactly

$$10v - (4v + 12) = 6v - 12$$

zero occurrences. The two empirical laws

$$\mu_{\mathsf{H},v} := \frac{1}{5}\sum_{r=1}^{5} \delta_{H_r}, \qquad \mu_{\mathsf{L},v} := \frac{1}{5}\sum_{r=1}^{5} \delta_{L_r}$$

therefore have exactly the same zero-inclusive intensity ν_v. At the known common dimension N_v, they consequently also have the same indexed Petz/down-arrow profile, pointwise at every positive order.

Exact equality of one-guess success. For $v > 2$, $a > b > t > 0$; at $v = 2$ all three values tie. The following top-one and top-two identities are unaffected by those ties. The unnormalized high maximum sum is

$$\sum_{r=1}^{5} F_1(H_r) = a + 4b = \frac{3}{2} - t, \tag{52}$$

whereas the low maximum sum is

$$\sum_{r=1}^{5} F_1(L_r) = 2a + 2b + t = \frac{3}{2} - t. \tag{53}$$

The final t is the maximum of the diffuse row L_5; retaining it is essential for exact equality at finite v. Dividing (52)–(53) by five gives the common value

$$r_v = \frac{1}{5}\left(\frac{3}{2} - t\right).$$

Exact two-guess gap. The unnormalized high top-two sum is

$$\sum_{r=1}^{5} F_2(H_r) = 2a + 4(2b) = 3 - 2t. \tag{54}$$

On the low side,

$$\begin{aligned} \sum_{r=1}^{5} F_2(L_r) &= 2(a + t) + 2(2b) + 2t \\ &= 2 + 2t. \end{aligned} \tag{55}$$

It follows that

$$\begin{aligned} R_2(\mu_{\mathsf{H},v}) - R_2(\mu_{\mathsf{L},v}) &= \frac{1}{5}(1 - 4t) \\ &= \frac{1}{5} - \frac{2}{5v}. \end{aligned} \tag{56}$$

At $v = 2$, both sides reduce to five copies of the uniform four-atom spectrum and the gap is zero. For $v > 2$, the displayed gap is positive and strictly below $1/5$, and it converges to $1/5$ as $v \to \infty$.

Finite common-uniform-prior realization. The five-row laws need not share a labeled barycenter. Coordinate symmetrization gives them a common uniform prior while preserving their intensities and list accuracies. Set $N = N_v$ and, for $\mathsf{E} \in \{\mathsf{H}, \mathsf{L}\}$, define

$$\widetilde{\mu}_{\mathsf{E},v} := \frac{1}{N!} \sum_{\pi \in S_N} \pi_{\#} \mu_{\mathsf{E},v}. \tag{57}$$

Every F_k and every separable coordinate sum is permutation-invariant. Therefore

$$R_k(\widetilde{\mu}_{\mathsf{E},v}) = R_k(\mu_{\mathsf{E},v}), \qquad \nu_{\widetilde{\mu}_{\mathsf{E},v}} = \nu_{\mu_{\mathsf{E},v}}. \tag{58}$$

More strongly, let $f : [0, 1] \to \mathbb{R}$ be bounded and Borel. For every coordinate j, each source coordinate appears in position j under exactly $(N - 1)!$ permutations, so

$$\begin{aligned} \int f(q_j)\, \widetilde{\mu}_{\mathsf{E},v}(\,\mathrm{d}q) &= \frac{1}{N!} \sum_{\pi \in S_N} \int f((\pi q)_j)\, \mu_{\mathsf{E},v}(\,\mathrm{d}q) \\ &= \frac{1}{N} \int \sum_{i=1}^{N} f(q_i)\, \mu_{\mathsf{E},v}(\,\mathrm{d}q) = \frac{1}{N} \int f \,\mathrm{d}\nu_v. \end{aligned} \tag{59}$$

Hence, for both $\mathsf{E} = \mathsf{H}$ and $\mathsf{E} = \mathsf{L}$,

$$\xi_{\widetilde{\mu}_{\mathsf{E},v},j} = \frac{1}{N}\nu_v, \qquad j = 1, \ldots, N. \tag{60}$$

Taking $f(x) = x$ gives

$$\int q_j \, \widetilde{\mu}_{\mathsf{E},v}(dq) = \frac{1}{N}. \tag{61}$$

Thus both symmetrized laws have identical complete labeled coordinate marginals and the same uniform barycenter.

For completeness, an explicit finite channel realizes each law. Retain the pair $s = (r, \pi) \in \{1, \dots, 5\} \times S_N$ as a labeled signal, define $q_r^{\mathsf{H}} := H_r$, $q_r^{\mathsf{L}} := L_r$, and let

$$q_{r,\pi}^{\mathsf{E}} := \pi q_r^{\mathsf{E}}.$$

Write $q_{r,i}^{\mathsf{E}}$ for the i-th coordinate of q_r^{E}. Give the latent state the common prior $p_j = 1/N$, and define

$$W_{\mathsf{E}}((r, \pi) \mid j) := \frac{N}{5N!} \, q_{r,\pi,j}^{\mathsf{E}}. \tag{62}$$

For each j,

$$\begin{aligned} \sum_{r=1}^{5} \sum_{\pi \in S_N} W_{\mathsf{E}}((r, \pi) \mid j) &= \frac{N}{5N!} \sum_{r=1}^{5} (N-1)! \sum_{i=1}^{N} q_{r,i}^{\mathsf{E}} \\ &= 1, \end{aligned}$$

so (62) is a channel. Its signal marginal is

$$\Pr_{\mathsf{E}}\{S = (r, \pi)\} = \sum_{j=1}^{N} \frac{1}{N} W_{\mathsf{E}}((r, \pi) \mid j) = \frac{1}{5N!},$$

and Bayes' rule gives

$$\Pr_{\mathsf{E}}\{X = j \mid S = (r, \pi)\} = q_{r,\pi,j}^{\mathsf{E}}.$$

The induced posterior laws are exactly $\widetilde{\mu}_{\mathsf{H},v}$ and $\widetilde{\mu}_{\mathsf{L},v}$. Both are finite and have the uniform prior on the same N_v labels. They match every labeled coordinate marginal, retain the common intensity and R_1, and preserve the gap in (56). These finite pairs approach the sharp constant; the construction makes no finite-attainment claim.

E General-Rank Constructions and Certificates

This section proves the remaining assertions of theorem 4.1. The quarter upper bound is proved in Section E.1 below. The top-2 result and its stronger minimax statement are proved in Sections B to D. All multiplicity exponents below mean repeated coordinates; rows are padded with zeros to their common stated dimension.

E.1 The universal quarter certificate

Lemma E.1 (Quarter residual bound)**.** *For integers $N \geq k \geq 1$, let $t = 1/(4k)$ and*

$$D_k(q) = \frac{1}{4} + \sum_{i=1}^{N} (q_i - t)_+ - F_k(q).$$

Then $0 \leq D_k(q) \leq 1/4$ for every $q \in \Delta_N$, including ties and zeros.

Proof. Sort the coordinates and put $m = \#\{i : q_i > t\}$. Since $kt = 1/4$, if $m \leq k$ then

$$D_k(q) = (k - m)t - \sum_{j=m+1}^{k} q_{(j)}.$$

Each term in this sum lies in $[0, t]$, so $0 \leq D_k(q) \leq (k - m)t \leq 1/4$. This includes $m = 0$ and $m = k$.

If $m > k$, cancellation instead gives

$$D_k(q) = \sum_{j=k+1}^{m} (q_{(j)} - t) \geq 0.$$

The average of the largest k entries is at least the average of the largest m entries. Hence, writing $S_m = \sum_{j=1}^{m} q_{(j)} \leq 1$,

$$\sum_{j=k+1}^{m} q_{(j)} \leq \frac{m-k}{m} S_m \leq \frac{m-k}{m}.$$

Consequently,

$$D_k(q) \leq (m-k)\left(\frac{1}{m} - \frac{1}{4k}\right) = \frac{1}{4} - \frac{(m-2k)^2}{4km} \leq \frac{1}{4}.$$

Coordinates equal to t contribute zero to the hinge sum and satisfy the same tail bounds, so the two cases also cover ties. □

For a posterior report s, the expectation of the hinge sum is $J_k(s)$. The lemma therefore gives

$$R_k(\mu) = J_k(s) + \frac{1}{4} - \mathbb{E}_\mu D_k(Q) \in [J_k(s), J_k(s) + 1/4].$$

Thus any two compatible Borel laws differ in R_k by at most $1/4$. For $N \geq 2$, the largest $k-1$ of the remaining $N-1$ coordinates have at least their proportional share of $1 - F_1(q)$, while each is at most $F_1(q)$. It follows that

$$F_1(q) + \frac{k-1}{N-1}(1 - F_1(q)) \leq F_k(q) \leq \min(kF_1(q), 1).$$

Taking expectations proves Equation (12). The finite family in Section E.3 then supplies the matching large-rank limit.

E.2 Nonidentification at every nontrivial rank in every dimension

First consider two three-coordinate trades. Each column in the following tables is a uniform probability law over its rows:

Divide every entry by 18		Divide every entry by 22	
T_2^+	T_2^-	T_3^+	T_3^-
$(1,7,10)$	$(1,4,13)$	$(1,8,13)$	$(1,4,17)$
$(1,8,9)$	$(1,5,12)$	$(1,9,12)$	$(1,5,16)$
$(2,3,13)$	$(2,6,10)$	$(1,10,11)$	$(1,6,15)$
$(2,4,12)$	$(2,7,9)$	$(2,3,17)$	$(2,7,13)$
$(5,6,7)$	$(3,7,8)$	$(2,4,16)$	$(2,8,12)$
		$(2,5,15)$	$(2,9,11)$
		$(6,7,9)$	$(3,9,10)$

Each trade preserves the pooled coordinate inventory. Direct summation, with Δ denoting the plus law minus the minus law, gives

$$\Delta_{T_2}(R_1, R_2) = -(1,2)/90, \qquad \Delta_{T_3}(R_1, R_2) = -(1,3)/154.$$

Set $t = 45/122$ and form

$$\mu_{\mathsf{A}} = tT_2^+ + (1-t)T_3^-, \qquad \mu_{\mathsf{B}} = tT_2^- + (1-t)T_3^+.$$

The intensities match, while $\Delta R_1 = -t/90 + (1-t)/154 = 0$ and $\Delta R_2 = -2t/90 + 3(1-t)/154 = 1/244$.

For $2 \leq k < N$, put $c = 1/(2k-3)$ and map each three-coordinate row to

$$T_{N,k}(q) = \left((2c)^{[k-2]}, cq_1, cq_2, cq_3, 0^{[N-k-1]}\right).$$

The coordinate sum is $2c(k-2)+c=1$, and every anchor exceeds every scaled base coordinate. At $k=2$ the map is zero padding; at $k\geq 3$ the largest coordinate is the constant $2c$. Hence top-1 still matches. Moreover,

$$F_k(T_{N,k}(q)) = 2c(k-2) + cF_2(q), \qquad \Delta R_k = \frac{1}{244(2k-3)} > 0.$$

Scaling and fixed padding preserve matching intensities. Averaging each law over the N cyclic coordinate shifts makes every labeled marginal ν/N and preserves every F_j, because these functions are permutation invariant. The common barycenter is $(1/N,\dots,1/N)$. The realization $Q\sim\mu$, $X\mid Q\sim Q$ gives two finite posterior experiments with that prior. This proves the fixed-dimension assertion, without claiming that the displayed small gap is extremal or positive for every report.

E.3 The finite family approaching the quarter bound

Fix $k\geq 2$ and choose $D>2k$ divisible by $4k$. Put

$$a=\frac{k+2}{4k}, \qquad b=\frac{1}{2k}, \qquad d=\frac{1}{D}.$$

The following five row types all sum to one and fit in dimension D:

$$U=(a^{[2]},b^{[k-2]}), \quad V=(b^{[k]},d^{[D/2]}),$$
$$W=(a,d^{[D(3k-2)/(4k)]}), \quad X=(b^{[2k]}), \quad Z=(d^{[D]}).$$

Take P copies of U and A copies of V on the high side, and R copies of W, S copies of X, and T copies of Z on the low side, where

$$P=k(D-2k), \qquad A=D(k^2+3k-2)-2k(3k-2),$$
$$R=2P, \qquad S=\frac{D(k^2+4k-4)-8k(k-1)}{2}, \qquad T=\frac{Dk^2}{2}.$$

These are positive integers. Both sides have row count

$$B_k(D)=P+A=R+S+T=D(k^2+4k-2)-4k(2k-1).$$

To verify the inventories, match each of the three positive atom sizes:

$$2P=R, \qquad (k-2)P+kA=2kS, \qquad AD/2=RD(3k-2)/(4k)+TD.$$

Equal row count and dimension also match the zeros after padding. The mean maximum agrees because

$$Pa+Ab=Ra+Sb+T/D.$$

The five top-k sums are, respectively,

$$1, \qquad \frac{1}{2}, \qquad a+\frac{k-1}{D}, \qquad \frac{1}{2}, \qquad \frac{k}{D}.$$

Subtracting their averages gives

$$\Delta R_k=\frac{(D-2k)[D(k^2+4k-4)-8k(k-1)]}{4D[D(k^2+4k-2)-4k(2k-1)]}. \tag{63}$$

Cyclic coordinate averaging again matches all labeled marginals and produces the uniform prior. Taking $D\to\infty$ along multiples of $4k$ gives

$$\lim_D \Delta R_k=\frac{k^2+4k-4}{4(k^2+4k-2)}=\frac{1}{4}-\frac{1}{2(k^2+4k-2)}.$$

The construction also works at $k=2$: U simply has no b coordinates. At $k=3$ the gap is

$$G_3(D)=\frac{(D-6)(17D-48)}{4D(19D-60)}, \qquad \frac{17}{76}-G_3(D)=\frac{3(305D-912)}{38D(19D-60)}>0.$$

Thus the following top-3 upper certificate is sharp as a global supremum.

E.4 A SHARP TOP-3 SCALAR CERTIFICATE

Let $\beta = 6/19$, $\lambda = 13/19$, $C = 51/19 = \lambda + 2$, and $a_0 = 5/34$. Define the continuous potential

$$\psi(t) = \begin{cases} \frac{51}{19}t^2, & 0 \le t \le \frac{5}{34}, \\ \frac{16}{19}t - \frac{5}{76}, & \frac{5}{34} \le t \le \frac{1}{6}, \\ \frac{17}{19}t - \frac{17}{228}, & \frac{1}{6} \le t \le \frac{1}{3}, \\ \frac{14}{19}t - \frac{5}{228}, & \frac{1}{3} \le t \le \frac{5}{12}, \\ \frac{13}{19}t, & \frac{5}{12} \le t \le 1. \end{cases} \tag{64}$$

We prove, for every finite probability row including ties and zeros,

$$0 \le h_3(q) := F_3(q) - \beta F_1(q) - \sum_i \psi(q_i) \le \frac{17}{76}. \tag{65}$$

At the four junctions $5/34, 1/6, 1/3, 5/12$, the potential values are $75/1292, 17/228, 17/76, 65/228$ and the ratios $\psi(t)/t$ are $15/38, 17/38, 51/76, 13/19$. The ratio is nonnegative and nondecreasing on all branches, and bounded by λ. The first three branches form a convex function on $[0, 1/3]$: the successive terminal or constant slopes are $15/19 < 16/19 < 17/19$. Put $A(t) = \lambda t - \psi(t)$ and $B(t) = t - \psi(t)$. Both are nonnegative, and B is strictly increasing: its branch derivatives have lower bounds $4/19, 3/19, 2/19, 5/19, 6/19$.

Lower bound for a small third coordinate. Write $x = q_{(1)}$, $y = q_{(2)}$, $z = q_{(3)}$, allowing missing entries to be zero. Suppose $0 \le z \le a_0$. The tail mass is $T = 1 - x - y - z$, and

$$\sum_{j \ge 4} \psi(q_{(j)}) = C \sum_{j \ge 4} q_{(j)}^2 \le CzT.$$

It follows that

$$h_3(q) \ge [A(x) + Czx] + [B(y) + Czy] + z - Cz.$$

For $t \ge z$, we claim

$$A(t) + Czt \ge \lambda z, \qquad B(t) + Czt \ge z. \tag{66}$$

For $t \le a_0$, the differences factor as $(t - z)(\lambda - Ct)$ and $(t - z)(1 - Ct)$, both nonnegative since $Ca_0 = 15/38 < \lambda < 1$. For $t \ge a_0$, they are affine in z, so it suffices to check $z = 0, a_0$. At zero they are $A(t), B(t) \ge 0$; at a_0 the differences are the following nonnegative affine functions:

t interval	$A(t) + Ca_0t - \lambda a_0$	$B(t) + Ca_0t - a_0$
$[5/34, 1/6]$	$9(34t - 5)/1292$	$21(34t - 5)/1292$
$[1/6, 1/3]$	$(714t - 101)/3876$	$(1938t - 281)/3876$
$[1/3, 5/12]$	$(1326t - 305)/3876$	$5(510t - 97)/3876$
$[5/12, 1]$	$5(51t - 13)/646$	$(459t - 95)/646$

Equation (66) yields $h_3 \ge \lambda z + z + z - Cz = 0$.

Lower bound for a large third coordinate. Suppose $z > a_0$. A convex function g with $g(0) = 0$, applied to entries in $[0, z]$ of total mass T, has maximum sum at $(z^{[n]}, r, 0, \ldots)$, where $n = \lfloor T/z \rfloor$ and $r = T - nz$. Indeed, moving mass from a smaller interior entry to a larger one until one becomes zero or reaches z cannot decrease a convex sum; iterate. Because $z \le 1/3$, this applies to ψ on the whole tail. Also $T/z \le 1/z - 3 < 19/5$, so $n \in \{0, 1, 2, 3\}$. Consequently

$$h_3(q) \ge D_n(x, y, z, r) := \lambda x + y + z - \psi(x) - \psi(y) - (n + 1)\psi(z) - \psi(r),$$

where

$$x \ge y \ge z \ge r \ge 0, \qquad x + y + (n + 1)z + r = 1.$$

Partition each variable at $0, 5/34, 1/6, 1/3, 5/12, 1$ and use closed cells with $z \ge a_0$. On every such cell, x, y, z lie on affine branches; $\psi(r)$ is affine or quadratic. After eliminating r, D_n is affine or concave, so its minimum is attained at a vertex. Exact substitution gives

n	0	1	2	3
Nonempty branch cells	26	18	11	4
Distinct vertices within n	36	28	16	7
$\min D_n$	$17/228$	$17/1140$	0	0

(67)

For clarity, this is a finite exact arithmetic step in the proof, not a sampled numerical bound. It can be reproduced as follows. For each n, intersect the normalization hyperplane with every independent triple of planes of the form $v = b$ (one of the four variables, one of the six breakpoints), $x = y$, $y = z$, or $z = r$. Retain points satisfying ordering, nonnegativity, and $z \geq a_0$, deduplicate, and evaluate Equation (64). These are all vertices of the subdivision: a vertex in the three-dimensional normalization hyperplane has three independent active branch or ordering boundaries. The supplied checker performs this enumeration by rational four-variable Gaussian elimination and lists all 87 evaluations. A separate branch-by-branch enumeration using three-variable determinants verifies the same sets and minima. Their nonnegativity completes the lower bound in every dimension.

Upper bound. Dropping the nonnegative tail potential gives

$$h_3(q) \leq U(x, y, z) := A(x) + B(y) + B(z).$$

For fixed x, z, increasing y increases U, so the relaxed optimum has $y = \min(x, 1 - x - z)$. If $x \leq (1 - z)/2$, then $y = x$. The derivative signs of $S(t) = A(t) + B(t)$ across the five potential branches are positive, zero, negative, positive, positive, respectively; on the last three branches the derivatives are $-2/19, 4/19, 6/19$. For $z \leq 1/6$, the maximum over $z \leq x \leq (1 - z)/2$ is therefore on the flat interval $[5/34, 1/6]$ or at $x = (1 - z)/2$. On the flat interval,

$$S(x) + B(z) \leq 5/38 + B(1/6) = 5/38 + 7/76 = 17/76.$$

For $z \geq 1/6$, only a decrease followed by an increase remains, so the maximum is at $x = z$ or $x = (1 - z)/2$; the first gives $17/76$ by substitution.

If $x \geq (1 - z)/2$, then $y = 1 - z - x$ and $x \geq 1/3$. The function A is nonincreasing on $[1/3, 1]$ (slope $-1/19$, then zero), while $B(1 - z - x)$ decreases with x. Thus this regime also maximizes at $x = y = (1 - z)/2$. For $0 \leq z \leq 5/34$, direct substitution gives

$$\frac{17}{76} - U((1 - z)/2, (1 - z)/2, z) = \frac{(1 - 6z)(5 - 34z)}{76} \geq 0.$$

For $5/34 \leq z \leq 1/6$ and $1/6 \leq z \leq 1/3$, substitution gives equality. This proves the upper half of Equation (65).

Integration and sharpness. For any Borel posterior law, the row identity integrates to

$$R_3(\mu) = \frac{6}{19} R_1(\mu) + \int \psi \, \mathrm{d}\nu_\mu + \mathbb{E}_\mu h_3.$$

Matching top-1 and pooled measure cancel the first two terms. Two residual expectations in $[0, 17/76]$ differ by at most $17/76$. This argument requires no finite-support or equal-weight assumption. The finite lower family in Section E.3, followed by cyclic averaging, approaches $17/76$ with all labeled marginals matched. Hence $E_3 = 17/76$.

Verification scope. The accompanying manuscript checker verifies 171 fixed-dimension/rank constructions, 84 compressed lower-family members, 62,483 rational row/rank hinge checks, and all finite steps of the top-3 reduction. The newly reconstructed vertex file records all 87 rational points and values. Finite grids are regression evidence; the arbitrary-dimensional claims rely on the tail reductions, the exhaustive vertex argument, and integration above.

F Finite Dimension and Longer-List Certification

F.1 Finite-dimensional lower envelope

Let

$$\mathcal{D}_N^\star := \sup_{(\nu, r) \text{ feasible at } N} \mathcal{D}_{N,\nu,r}.$$

Corollary F.1 (Finite-dimensional lower envelope)**.** *For every $N \geq 4$,*

$$\boxed{\frac{1}{5} - \frac{2}{5\lfloor N/2 \rfloor} \leq \mathcal{D}_N^\star \leq \frac{1}{5}.} \tag{68}$$

For every $N \geq 6$, the positive lower bound occurs in an explicit slice and, after symmetrization, under a common uniform prior with every labeled coordinate marginal matched across the pair.

Table 5: **Certified gaps at familiar label dimensions.** Lower and upper bounds for maximal continuum top-2 ambiguity $\mathcal{D}_N^\star$, rounded outward. Lower entries through $N = 10$ use certified $1/48$-grid optima; $N = 20, 100$ use the analytic construction. Upper entries use the continuum cell-vertex certificates in Section F.2. Output budgets are unrestricted.

N	3	4	5	6	8	10	20	100
$\mathcal{D}_N^\star \geq$	0.0491	0.0952	0.1153	0.1261	0.1379	0.1458	0.1600	0.1920
$\mathcal{D}_N^\star \leq$	0.0625	0.1191	0.1320	0.1400	0.1500	0.1556	0.1723	0.1948

Proof of corollary F.1. Fix $N \geq 4$, set $v = \lfloor N/2 \rfloor$ and $d = 2v$, and take the five-row pair from Equation (49) in dimension d. Append $N - d$ zeros to every row. If $\iota_{d,N} : \Delta_d \to \Delta_N$ is this embedding, then

$$R_j((\iota_{d,N})_{\#}\mu) = R_j(\mu) \quad (j = 1, 2), \qquad \nu_{(\iota_{d,N})_{\#}\mu} = \nu_\mu + (N - d)\delta_0.$$

Consequently the padded laws retain equal intensity and equal R_1, while their R_2 gap remains

$$\frac{1}{5} - \frac{2}{5v} = \frac{1}{5} - \frac{2}{5\lfloor N/2 \rfloor}.$$

More explicitly, with $t = 1/(2v)$, $a = 1/2 - t$, and $b = 1/4$, their common intensity and one-guess success are

$$\nu_N = \frac{1}{5}\left(2\delta_a + 8\delta_b + (4v + 2)\delta_t + (5N - 4v - 12)\delta_0\right), \qquad r_N = \frac{1}{5}\left(\frac{3}{2} - t\right).$$

Coincident displayed atoms are combined. This explicit slice proves the lower bound; the upper bound is theorem B.1 after taking the supremum over feasible (ν, r). Symmetrizing both padded laws over all N coordinates preserves their intensity and every R_j, gives both barycenter $(1/N, \ldots, 1/N)$, and hence supplies common-uniform-prior channels by the kernel construction in Section D. That construction applies to any finite symmetrized posterior law, including this padded pair. □

F.2 Continuum upper certificates at fixed dimension

Finite grids provide lower bounds on $\mathcal{D}_N^\star$. To certify an upper bound over all posterior rows, we instead use a piecewise-affine potential and a finite set that contains every vertex of its affine cells. Fix knots $0 = \tau_0 < \cdots < \tau_J = 1$, and let η be continuous and affine between consecutive knots. Write

$$u(q) = F_2(q) - cF_1(q) - \sum_i \eta(q_i).$$

Let $V_{N,\tau}$ consist of all sorted simplex rows whose coordinates are knots except for at most one distinct non-knot value, which may be repeated any number of times. This is finite: if the knot coordinates sum to s and the non-knot value occurs r times, its value is $(1 - s)/r$. Rows containing only knots are included.

Proposition F.2 (Finite vertex certificate for the continuum)**.** *For every finite $N \geq 2$, every such potential, and every $c \in \mathbb{R}$,*

$$\mathcal{D}_N^\star \leq \max_{q \in V_{N,\tau}} u(q) - \min_{q \in V_{N,\tau}} u(q). \tag{69}$$

The displayed extrema equal the extrema on the whole simplex. The bound also holds for matched labeled marginals, with unrestricted output budgets.

Proof. The residual is invariant under coordinate permutations, so restrict to $q_1 \geq \cdots \geq q_N$. Intersect this ordered simplex with all the coordinate knot intervals. These finitely many closed polytopes cover it, and u is affine on each: the first two order statistics are the first two coordinates, and every potential branch is affine. An extremum on each nonempty polytope occurs at a vertex.

At a vertex, group equal coordinates into maximal consecutive blocks. Every block at a knot is fixed. Suppose two distinct blocks have values strictly between knots, with multiplicities r and s. Increase the first block's value by ε/r and decrease the second's by ε/s, or make the opposite changes. For

Table 6: **Rational continuum upper certificates.** All displayed coefficients and residual endpoints are multiplied by D. The potential is affine at knots $0, 1/8, 1/4, 1/2, 3/4, 1$, with endpoint values zero. Dividing the last two columns by D gives its exact residual range; the difference bounds $\mathcal{D}^\star_N$.

N	D	Dc	$D\eta_{1/8}$	$D\eta_{1/4}$	$D\eta_{1/2}$	$D\eta_{3/4}$	$D\ell$	Du
3	16	8	-1	-1	2	2	8	9
4	336	96	-25	-34	20	18	240	280
5	1152	384	-61	-58	76	90	680	832
6	400	160	-19	-14	28	30	208	264
7	384	128	-21	-18	20	26	224	280
8	1120	320	-67	-62	44	70	704	872
9	2560	640	-163	-158	76	150	1696	2088
10	90	20	-7	-8	-2	-1	70	84
20	1672	464	-115	-86	44	58	1064	1352
100	48312	18224	-3663	-2622	-148	-2230	30088	39496

sufficiently small $\varepsilon > 0$, both perturbations preserve the strict gaps between distinct blocks, all knot interval constraints, and the total sum. The point is then the midpoint of two different feasible points, contradicting extremality. Thus at most one distinct block is off-knot, and every cell vertex lies in $V_{N,\tau}$. Since the latter set is itself contained in the simplex, its extrema are exactly the continuum extrema. Applying proposition B.2 proves the bound. □

Explicit certificates. We fix the same knots $\tau = (0, 1/8, 1/4, 1/2, 3/4, 1)$ for every reported dimension. An LP chooses c, the knot values, and a residual enclosure from the finite vertex inequalities. Subtracting an affine function from the potential changes the residual by a constant at fixed N, so impose $\eta(0) = \eta(1) = 0$. After rational reconstruction, an independent checker verifies every inequality on $V_{N,\tau}$ exactly. These checks certify the upper bound regardless of whether the numerical solver found an optimal potential. Table 6 gives all retained potentials. For example, at $N = 10$, take $c = 2/9$ and interior knot values

$$\big(\eta(1/8), \eta(1/4), \eta(1/2), \eta(3/4)\big) = \left(-\frac{7}{90}, -\frac{4}{45}, -\frac{1}{45}, -\frac{1}{90}\right).$$

The exact residual range is $[7/9, 14/15]$, giving $\mathcal{D}^\star_{10} \le 7/45 \simeq 0.155556$. Together with the existing $7/48$ lower witness, this brackets the ten-class continuum supremum in $[7/48, 7/45]$. At one hundred classes the bracket is $[24/125, 392/2013]$.

The independent checker enumerates a superset of the cell vertices: it chooses every multiset of positive knots with sum at most one, then fills the remaining mass with one repeated value and zeros. These knots allow at most eight positive knot coordinates. The generator and checker use different enumerations; exact checks include ties, knots, and zero padding. By proposition F.2, this finite coverage certifies a continuum bound. Every bound in the table is strictly below $1/5$, while equality with the corresponding fixed-dimensional supremum remains unproved.

F.3 Certified finite-grid optimization

At small N, grid-supported laws can improve the five-row lower bound. We optimize over a finite restriction of the population problem, allowing arbitrary output weights and optimizing the common summary itself. Let

$$\mathcal{P}_{N,M} := \{s \in \mathbb{Z}^N_{\ge 0} : s_1 \ge \cdots \ge s_N,\ \textstyle\sum_i s_i = M\}, \qquad c_j(s) := \#\{i : s_i = j\},$$

for $j = 0, \dots, M$. Each s represents the posterior spectrum s/M. Sorting preserves the pooled intensity and every list value. Conversely, uniformly applying its N cyclic coordinate shifts gives each labeled coordinate the same marginal ν/N and gives the uniform class prior. Thus no restriction on the grid optimum is introduced by using sorted spectra: any grid-supported pair reduces to these weights, and every feasible pair of weights has a common-uniform-prior realization with matched labeled marginals.

Table 7: **Exact finite-grid optima for three to ten classes.** Every grid entry has rational primal and dual certificates. Nested denominators $12, 24, 48$ enlarge the feasible grid; their values are lower bounds for $\mathcal{D}_N^\star$, with no claimed continuum discretization-error bound. The five-row column evaluates Equation (68).

N	Five-row bound	$G_{N,12}$	$G_{N,24}$	$G_{N,48}$
3	—	$7/156$	$3/62$	$3/61$
4	0	$2/21$	$2/21$	$2/21$
5	0	$1/9$	$11/96$	$3/26$
6	$1/15$	$1/8$	$265/2112$	$55/436$
7	$1/15$	$1/8$	$15/112$	$17/126$
8	$1/10$	$1/8$	$50/363$	$1775/12864$
9	$1/10$	$1/8$	$183/1280$	$28/195$
10	$3/25$	$1/8$	$7/48$	$7/48$

For probability weights a_s, b_s, solve

$$\begin{aligned} G_{N,M} := \max_{a,b\geq 0} \quad & \frac{1}{M}\sum_{s\in\mathcal{P}_{N,M}}(a_s-b_s)(s_1+s_2) \\ \text{subject to} \quad & \sum_s a_s = \sum_s b_s = 1, \\ & \sum_s (a_s-b_s)c_j(s) = 0 \quad (j=0,\dots,M), \\ & \sum_s (a_s-b_s)s_1 = 0. \end{aligned} \tag{70}$$

The constraints match the complete pooled intensity, including zero atoms, and the top-1 value. They optimize over feasible summaries rather than impose a particular report. Consequently $G_{N,M} \leq \mathcal{D}_N^\star$, and the lower bound is realized by finite laws. Matching the labeled marginals follows from the symmetrization just described, not from assuming that an arbitrary sorted pair already matches its labeled columns.

Exact certificates. Floating-point linear programming is used only to discover candidate weights and dual variables. Every retained result is verified with rational arithmetic. A dual certificate consists of $u, v, \beta, z_0, \dots, z_M$ such that, for every $s \in \mathcal{P}_{N,M}$,

$$u + \sum_{j=0}^{M} z_j c_j(s) + \beta s_1 \geq s_1 + s_2, \qquad v - \sum_{j=0}^{M} z_j c_j(s) - \beta s_1 \geq -(s_1+s_2). \tag{71}$$

Multiply the first inequality by a_s, the second by b_s, and sum. The matched terms cancel, giving $G_{N,M} \leq (u+v)/M$. Our saved feasible weights attain this rational upper bound exactly. An independent verifier enumerates all spectra, checks every inequality and primal equality, and compares the two objectives without numerical tolerances. It also explicitly forms the cyclic realizations and verifies every labeled marginal, uniform prior, Bayes-kernel normalization, and preserved list values. A separate integer partition count checks enumeration coverage; deliberate primal and dual corruptions are rejected.

Results and limits. We certify all 24 problems at $N = 3, \dots, 10$ and $M \in \{12, 24, 48\}$, covering 142,777 spectrum cases and 285,554 dual inequalities. Table 7 reports exact rational values. At $M = 48$, the certified gap is $3/61$ for three classes, $3/26$ for five, and $7/48$ for ten; all eight dimensions improve the five-row lower envelope where it applies. The nested grids give monotone lower bounds. Agreement at two or more grid sizes does not certify a continuum optimum: rounding general posterior rows can destroy exact marginal and top-1 matching, and we do not supply a discretization-error bound. The grid dual variables certify only grid optimality. Upper bounds for all real-valued posterior rows require a separate continuum argument, such as the cell-vertex certificates in Section F.2.

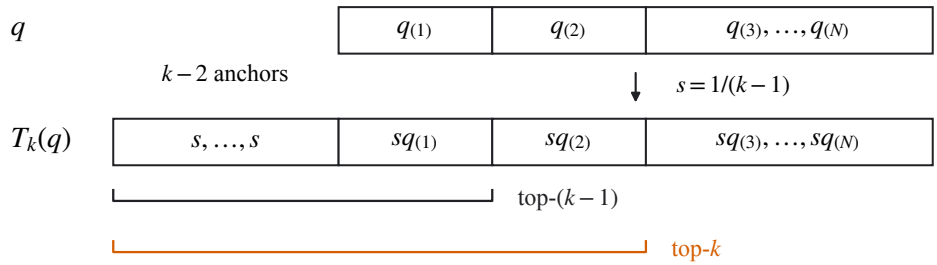


Figure 6: **Anchor lift to longer lists.** The map T_k inserts $k-2$ anchors of size $s = 1/(k-1)$ before a scaled posterior row. Hence the first $k-1$ list positions depend only on the anchors and the base maximum, while slot k adds the base runner-up. The brackets mark these two lists; box widths do not encode probability. Matched base summaries therefore preserve all earlier R_j, $j < k$, with $\Delta R_k = \Delta R_2/(k-1)$. The exact constant $1/[5(k-1)]$ applies only to this structured lifted family.

The discovery script `code/finite_grid_lp.py` uses SciPy's HiGHS interface. Rational weights and dual variables are saved in `code/finite_grid_lp_certificates.json`, and `code/check_finite_grid_lp.py` verifies them using only Python's standard library. Solver versions, exact check results, and table inputs accompany these artifacts. All values concern finite population laws.

F.4 Anchor identities and budget lifting

For $k \geq 2$, put $s_k = (k-1)^{-1}$ and define

$$T_k(q) = \big(s_k^{[k-2]}, s_k q_1, \ldots, s_k q_N\big) \in \Delta_{N+k-2}. \tag{72}$$

Let $\mathfrak{A}_k$ be the supremum of the R_k gap over lifted pairs whose base laws have equal intensity and R_1.

Proposition F.3 (Anchor-lifted non-identifiability)**.** *For every $k \geq 2$,*

$$\boxed{\mathfrak{A}_k = \frac{1}{5(k-1)}.} \tag{73}$$

Every admissible lifted pair has equal complete profiles and equal R_j for all $1 \leq j < k$; after symmetrization, the sharp sequence also matches every labeled coordinate marginal and has a common uniform prior.

Hence all labeled coordinate marginals and all earlier Bayes list accuracies $R_1, \ldots, R_{k-1}$ need not determine R_k. The constant is sharp within the image of T_k and supplies a lower bound for the unrestricted problem.

For $s = s_k$, let $S_s(x) = sx$.

Lemma F.4 (Anchor identities)**.** *For every $q \in \Delta_N$,*

$$F_j(T_k(q)) = js, \qquad 1 \leq j \leq k-2, \tag{74}$$

$$F_{k-1}(T_k(q)) = s\big(k-2+F_1(q)\big), \tag{75}$$

$$F_k(T_k(q)) = s\big(k-2+F_2(q)\big). \tag{76}$$

For $\mu \in \mathcal{P}(\Delta_N)$,

$$\nu_{(T_k)_{\#}\mu} = (k-2)\delta_s + (S_s)_{\#}\nu_\mu, \tag{77}$$

$$M_\alpha((T_k)_{\#}\mu) = s^\alpha\big(k-2+M_\alpha(\mu)\big), \qquad \alpha > 0. \tag{78}$$

If $p = \int q\,\mu(\mathrm{d}q)$, the lifted barycenter is

$$p^{\uparrow} = \big(s^{[k-2]}, sp_1, \ldots, sp_N\big). \tag{79}$$

Proof. Normalization follows from $(k-1)s = 1$. Each scaled original coordinate satisfies $sq_i \le s$, so the first $k-2$ order statistics may be taken to be the anchors, followed by $sq_{(1)}$ and $sq_{(2)}$. Ties do not affect the sums. This proves the three F_j identities. Each lifted row contributes $k-2$ atoms of size s, while an original atom x becomes sx; integration proves Equations (77) and (78). Coordinatewise integration gives Equation (79). □

At the Shannon endpoint, direct calculation gives

$$H_1^{\downarrow}((T_k)_{\#}\mu) = \log(k-1) + sH_1^{\downarrow}(\mu). \tag{80}$$

Thus equality of the entire indexed profile is preserved.

Proof of proposition F.3. Let the base laws have equal intensity and equal R_1. Equation (77) gives equal lifted intensities and hence equal complete profiles. Equations (74)–(75) give equality of every R_j with $j < k$, while (76) gives

$$R_k((T_k)_{\#}\mu_{\mathsf{A}}) - R_k((T_k)_{\#}\mu_{\mathsf{B}}) = \frac{R_2(\mu_{\mathsf{A}}) - R_2(\mu_{\mathsf{B}})}{k-1}. \tag{81}$$

The fixed-slice theorem yields $\mathfrak{A}_k \le 1/[5(k-1)]$. Conversely, apply T_k to the five-row pair at parameter v; by Equations (56) and (81), its lifted gap is exactly

$$\frac{1}{k-1}\left(\frac{1}{5} - \frac{2}{5v}\right).$$

Letting $v \to \infty$ proves the reverse inequality at the level of the supremum.

For the operational statement, first symmetrize each base five-row law, giving both the uniform prior $p = (1/N, \dots, 1/N)$. Their lifts share the prior

$$p^{\uparrow} = \left(s^{[k-2]}, (s/N)^{[N]}\right)$$

by Equation (79). Symmetrizing the lifted laws over all $N + k - 2$ coordinates, if desired, preserves every symmetric quantity and makes the common prior uniform. □

Within the image of T_k, the matched summaries can also be recovered from the lift. Subtract $(k-2)\delta_s$ from Equation (77) and invert S_s to recover the base intensity; use Equation (75) to recover base R_1 from lifted R_{k-1}. Thus the lifted and base matching conditions are equivalent, establishing the exact diameter of this structured subclass.

BLOCK II: REPORT-SPECIFIC OPTIMIZATION

G REPORT-SPECIFIC CERTIFICATES AND A COUPLING-THEORY COMPARISON

G.1 ENDPOINT EXISTENCE, DUALITY, AND CONDITIONAL MINIMAX ERROR

We prove proposition 4.2 for finite $N \geq 1$, $1 \leq k \leq N$, and a feasible report with Borel probability marginals ξ_i on $[0,1]$. No atom, density, positive-prior, or interior-feasibility assumption is used.

Primal compactness and the full interval. The simplex is compact and metrizable, so $\mathcal{P}(\Delta_N)$ is compact under weak convergence. A coordinate projection is continuous; hence the pushforward map $\mu \mapsto (\mathrm{pr}_i)_{\#}\mu$ is weakly continuous. The maps F_1 and F_k are continuous, since each is the maximum of finitely many coordinate sums. Thus every defining constraint of $\mathcal{C}(s)$ is closed. The set is nonempty by feasibility, compact, and convex because its constraints are affine in the measure. The continuous affine map $\mu \mapsto \int F_k \, \mathrm{d}\mu$ attains its minimum and maximum. If μ_- and μ_+ attain them, then $(1-t)\mu_- + t\mu_+$, $0 \leq t \leq 1$, remains in $\mathcal{C}(s)$ and has value $(1-t)L_k(s) + tU_k(s)$. This proves both the interval claim and the diameter identity, including the case $L_k(s) = U_k(s)$.

Specializing constrained transport duality. Work first on the compact cube $D = [0,1]^N$. Let $\Pi(\boldsymbol{\xi})$ be its couplings with coordinate marginals ξ_i, and put

$$d(q) = \left| \sum_i q_i - 1 \right|, \qquad g(q) = F_1(q) - r, \qquad W = \mathrm{span}\{d, g\}.$$

All these functions and the costs $\pm F_k$ are continuous and bounded. A coupling annihilates W if and only if it has $\int g \, \mathrm{d}\mu = 0$ and $\int d \, \mathrm{d}\mu = 0$. Since $d \geq 0$ and its zero set is Δ_N, the second condition is equivalent to simplex support. The constrained couplings are therefore exactly $\mathcal{C}(s)$. The spaces are Polish, the constraints and costs lie in the continuous integrable class, and the feasible set is nonempty, so Zaev (2015, Theorem 2.1) applies.

For the cost F_k, its dual yields the supremum of $\sum_i \int f_i \, \mathrm{d}\xi_i$ over continuous f_i and scalars β, γ such that

$$\sum_i f_i(q_i) + \beta(F_1(q) - r) + \gamma d(q) \leq F_k(q) \quad (q \in D).$$

On Δ_N, this is a lower envelope G in Equation (15), with $a = -\beta r$ and the same objective. Consequently the supremum in Equation (15) is at least $L_k(s)$. Conversely, integrating any of its lower envelopes against a feasible posterior law gives $J_s(G) \leq L_k(s)$. Both inequalities prove equality. Applying the same argument to $-F_k$, and negating the resulting lower envelopes, proves the upper equality. The argument requires neither extension of an arbitrary simplex envelope to the cube nor attainment of the dual optimum.

The conditional decision problem. The report is fixed, so a possibly randomized estimator has one output law under both endpoint posterior laws. If its output is T, then

$$\mathbb{E}|T - L_k(s)| + \mathbb{E}|T - U_k(s)| \geq U_k(s) - L_k(s).$$

One endpoint loss is at least half the width. The deterministic midpoint has error at most half the width for every value in the interval, and thus attains this conditional minimax value. For $k = 1$, the interval is $\{r\}$; for $k = N$, it is $\{1\}$, including $N = 1$. Ties and zero-prior coordinates cause no discontinuity or additional restriction.

G.2 WHY FINITE MARGINAL SUPPORTS GIVE EXACT CONTINUUM ENDPOINTS

Let $A_i = \mathrm{supp}\, \xi_i$ be finite. For each compatible law, $\mu\{q : q_i \notin A_i\} = 0$. A finite union over coordinates shows that its support lies in $K = (\prod_i A_i) \cap \Delta_N$. Conversely, nonnegative weights on K satisfying the report equations define a compatible posterior law. Explicitly, the entire fixed-report continuum problem is the LP

$$\min_{w \geq 0} / \max_{w \geq 0} \sum_{q \in K} w_q F_k(q), \qquad \sum_q w_q = 1, \quad \sum_q w_q F_1(q) = r, \quad \sum_{q : q_i = x} w_q = \xi_i(\{x\}) \quad (x \in A_i).$$

Table 8: **Exact intervals for specified posterior reports.** The last column is the width of the universal interval after elementary rank bounds; it need not be attainable. All endpoints and widths are exact. Inputs and witnesses are specified in Section G.2.

Report	N	r	$[L(s), U(s)]$	Width	Clipped bound
Identified	4	$9/20$	$[29/40, 29/40]$	0	$23/120$
Half toy + vertex	4	$3/4$	$[7/8, 11/12]$	$1/24$	$19/180$
Figure 3	4	$1/2$	$[3/4, 5/6]$	$1/12$	$1/5$
Five-class report	5	$151/312$	$[161/234, 94/117]$	$3/26$	$1/5$

Its lower and upper duals assign an independent value $f_i(x)$ to each labeled support atom, together with normalization and top-1 multipliers a, β. The respective inequalities

$$a + \beta F_1(q) + \sum_i f_i(q_i) \le F_k(q) \quad \text{or} \quad a + \beta F_1(q) + \sum_i f_i(q_i) \ge F_k(q) \qquad (q \in K)$$

give outer bounds by integration. A feasible law with the same value certifies an attained endpoint. A feasible bounded rational LP has rational optimal primal and dual solutions; finite-dimensional LP duality also gives attainment here. Redundant marginal and normalization equations do not invalidate either certificate.

The implementation uses HiGHS to discover solutions and reconstructs rational candidates. A separate standard-library checker accepts a candidate only after verifying normalization, every labeled marginal, top-1, the objective, and every dual inequality exactly. Discovery enumerates tuples recursively using bounds on remaining coordinate sums. Verification instead constructs the full Cartesian product and filters by exact sum one, with a default budget of 10^6 tuples. Both procedures retain class identities and allow nonexchangeable marginals. An unjustified restriction of K would certify only witness gaps, so the exhaustive domain check is essential to the reported endpoints.

The entire coverage curve. The image $\{(R_1(\mu), \dots, R_N(\mu)) : \mu \in \mathcal{C}(s)\}$ is compact and convex. Every one of its curves is nondecreasing and discretely concave, since $R_{k+1} - R_k = \mathbb{E}Q_{(k+1)}$ decreases with k. Its separate coordinate envelopes are not a Cartesian-product description of that image. The proof of corollary 4.3 uses two individually attained endpoints and therefore does not require the envelopes to be jointly feasible. For finite score/label laws or integer batches the same proof needs only monotonicity and endpoint attainment, not concavity or convexity of the feasible set.

Worked top-2 reports. The following examples retain $L = L_2$ and $U = U_2$.

The identified report is generated by equal masses at

$$(1/2, 1/4, 1/8, 1/8), \qquad (1/5, 3/10, 1/10, 2/5).$$

Among the sixteen Cartesian tuples from its labeled supports, only these two sum to one. Their marginal masses force both weights to $1/2$. The interval is consequently $\{29/40\}$, despite nonconstant marginals. The second report is the mixture $\frac{1}{2}\mu_{\mathsf{A}} + \frac{1}{2}\delta_{(0,0,0,1)}$, where μ_{A} is the Figure 3 law. Its exact interval is $[7/8, 11/12]$; the third report is Figure 3 itself, with interval $[3/4, 5/6]$.

The five-class report fixes the labeled marginals of the $N = 5, M = 48$ grid example after cyclic coordinate averaging, together with $r = 151/312$. Each coordinate has support

$$A = \{0, 3, 4, 6, 12, 15, 16, 18, 24, 32, 36\}/48.$$

Its $11^5 = 161{,}051$ Cartesian tuples contain 1,650 simplex rows. Exact endpoint certificates give $[161/234, 94/117]$, of width $3/26$. This certifies the full interval for the supplied marginal tuple. The supremum over all five-class marginal tuples remains a separate problem.

Baseline and verification scope. Write $B(s) = \frac{2}{5}r + \sum_i \int \phi \, \mathrm{d}\xi_i$. Table 8 intersects $[B(s), B(s) + 1/5]$ with

$$\left[r + \frac{1-r}{N-1}, \ \min(2r, 1) \right] \qquad (N \ge 2).$$

The checker covers $N = 1$, binary scores, zero-prior classes, and exact ties. Across eight reports it checks 161,129 Cartesian tuples, 1,668 feasible simplex rows, and 3,336 endpoint dual inequalities, rejecting deliberately altered inputs.

Exact empirical reports can reconstruct a generic batch. The preceding finite-support formulation also explains why removing explicit row identifiers need not prevent recovery of cross-class row associations.

Proposition G.1 (Generic empirical reconstruction)**.** *Let $N \geq 2$, $M \geq 1$, and let $S^{(1)}, \ldots, S^{(M)}$ be iid draws from a law absolutely continuous with respect to relative Lebesgue measure on Δ_N. Set $A_i = \{S_i^{(m)} : 1 \leq m \leq M\}$. Almost surely, each A_i has M distinct atoms and*

$$\left(\prod_{i=1}^{N} A_i\right) \cap \Delta_N = \{S^{(1)}, \ldots, S^{(M)}\}.$$

For arbitrary realized labels $X^{(m)} \in [N]$, the exact empirical laws

$$\widehat{\gamma}_i = \frac{1}{M} \sum_{m=1}^{M} \delta_{(S_i^{(m)}, \mathbf{1}\{X^{(m)}=i\})}$$

therefore identify the joint empirical score/label law, including each row's label and mass. In particular, they determine every empirical top-k hit rate under any fixed tie rule; top-1 accuracy adds no identifying information.

Proof. The joint score law is absolutely continuous on the product of simplices. For a nonconstant selection $(m_1, \ldots, m_N) \in [M]^N$, choose a row index m for which $I = \{i : m_i = m\}$ is nonempty and proper. In the affine hull of that row's simplex, changing coordinate $i \in I$ by t and coordinate $j \notin I$ by $-t$ changes $\sum_i S_i^{(m_i)}$ by t. Thus the equation $\sum_i S_i^{(m_i)} = 1$ defines a proper affine hyperplane of the product affine hull and has probability zero. A finite union over mixed selections leaves only the observed rows. Likewise, $S_i^{(m)} = S_i^{(n)}$ for $m \neq n$ is a proper affine hyperplane, so every coordinate's atoms are distinct almost surely. Each reconstructed row is consequently the only row containing any one of its coordinate atoms. The mass and binary tag at that atom in $\widehat{\gamma}_i$ fix the row's mass and its indicator of label i. Collecting the tags recovers its label and hence all top-k hits. □

Score marginals alone recover the empirical score law and its averages of F_k; realized label hits require the indicator tags. The posterior identity Equation (4) is a population identity and does not equate empirical score marginals with realized empirical score/label laws. This proposition asserts identification, without an efficiency or numerical stability guarantee. Its sampling assumption need not hold for floating-point neural outputs or finite image datasets, and continuous one-dimensional marginals alone do not suffice.

G.3 A WORKED COMPARISON WITH DEPENDENT-ORDER-STATISTIC THEORY

Rychlik (1993, Lemma 1) characterizes compatible marginal distributions of all order statistics for a prescribed common coordinate distribution. We use its explicit reproduction in Jaworski & Rychlik (2008, Theorem 1): for uniform coordinates, the rank CDFs must be ordered and sum to Nt. The characterization constrains rank marginals, not the almost-sure sum of coordinates. The following finite-support specialization makes the comparison exact; it includes the additional expected-maximum constraint from our problem.

Let $x_1 < \cdots < x_m$ be the atoms of a common coordinate law ξ, and let $p_{j\ell}$ be the probability that the jth *largest* coordinate equals x_ℓ. The exact rank-marginal relaxation is

$$\begin{aligned} &p_{j\ell} \geq 0, \qquad \sum_{\ell} p_{j\ell} = 1, \qquad \sum_{j} p_{j\ell} = N\xi(\{x_\ell\}), \\ &\sum_{\ell \leq h} p_{j\ell} \leq \sum_{\ell \leq h} p_{j+1,\ell} \qquad (j < N,\ h < m), \quad \sum_{\ell} x_\ell p_{1\ell} = r. \end{aligned} \tag{82}$$

Table 9: **What the closest frameworks supply.** The numerical relaxation is our exact specialization of the cited characterization, not a previously published numerical bound.

Result	Constraints represented	Consequence here
Rychlik rank characterization	Common marginals; we add the mean maximum	Exact relaxation without simplex support; width 0.24 for the specified $N = 16$ report.
Zaev, Theorem 2.1	Marginals and linear constraints on the law	Encodes the full problem and gives a variational dual; a potential still needs to be evaluated.
This paper's calculation	Simplex support, marginals, and mean maximum	Sharp $1/5$ and $17/76$ certificates at ranks two and three; general-rank bounds approaching $1/4$; exact report-specific intervals and extremal budgets.

Its objective is $\sum_\ell x_\ell(p_{1\ell} + p_{2\ell})$. These constraints are necessary for any coupling on $[0,1]^N$. They are sufficient as well: use a single uniform variable to couple the inverse CDFs of the rank marginals. The CDF inequalities order the resulting coordinates pointwise. A uniform coordinate permutation then gives each labeled coordinate the marginal ξ, by the pooling equality, and preserves the maximum and top-two sum. This proof also handles atoms and ties. The missing condition is $\sum_i q_i = 1$ on every row; the specified marginal means impose only its expectation.

The sharpness-family report at $N = 16$. At $v = 8$, the symmetrized five-row construction has

$$\xi = \frac{9}{20}\delta_0 + \frac{17}{40}\delta_{1/16} + \frac{1}{10}\delta_{1/4} + \frac{1}{40}\delta_{7/16}, \qquad r = \frac{23}{80}.$$

Optimizing Equation (82) gives exactly

$$\left[\frac{67}{200}, \frac{23}{40}\right], \qquad \text{width } \frac{6}{25} = 0.24. \tag{83}$$

The elementary expectation bounds are tight in the relaxation: $r + (1 - r)/15 = 67/200$ and $2r = 23/40$. The simplex-supported high law in Equation (49), after cyclic coordinate averaging, attains $23/40$, so the posterior upper endpoint is also $23/40$. The corresponding low law attains $17/40$, giving a posterior width of at least $3/20$; the universal bound gives at most $1/5$. These witnesses do not determine the posterior lower endpoint. In particular, the relaxed lower endpoint $67/200$ is strictly too low: the posterior lower endpoint is at least $23/40 - 1/5 = 3/8$. The checker also constructs particular relaxed endpoint witnesses outside the simplex. This describes those witnesses, not every optimizer; the high posterior law is an upper optimizer supported entirely on the simplex. Thus the rank-marginal characterization, even augmented with exact top-1, permits width $6/25$ and cannot alone imply the posterior width bound $1/5$.

For the fixed five-class report in Table 8, the rank relaxation is $[255/416, 94/117]$, of width $713/3744$, whereas the posterior interval is $[161/234, 94/117]$, of width $3/26$. Again the upper endpoints agree, while the relaxed lower endpoint is too low. Both relaxed endpoint laws, all rank-LP dual inequalities, and the simplex-supported upper witnesses are checked exactly. These are our evaluations of specific relaxations derived from the cited framework.

BLOCK III: EMPIRICAL PROTOCOLS AND CERTIFICATES

H CLASSWISE REPORTS ACROSS DATASETS, MODELS, AND SEEDS

Prespecified design. The balanced design comprises five datasets, three classifiers, and ten seeds per dataset/model pair. The datasets are Iris (Fisher, 1936), Wine (Aeberhard & Forina, 1992), Digits (Alpaydin & Kaynak, 1998; Scikit-learn developers, 2024), Dry Bean (UCI Machine Learning Repository, 2020), and Statlog Landsat Satellite (Srinivasan, 1993). Iris, Wine, and Digits are snapshotted from scikit-learn 1.5.2; the remaining datasets are downloaded from their UCI archives with recorded checksums. All original features and classes are used. Labels are indexed in sorted order. The Landsat source files are concatenated before splitting, so results use our repeated holdout protocol rather than the original benchmark split.

Before splitting, we remove exact duplicate feature vectors (one Iris, 68 Dry Bean); no conflicting-label groups occur. Table 10 gives all sample sizes. Seeds 20260911 through 20260920 specify ten stratified 70/30 splits per dataset, shared across models. All model parameters, score precision, and seeds are fixed before execution. Every declared cell is retained, including zero-width reports.

Models and scores. We use scikit-learn 1.8.0 implementations of `RandomForestClassifier` (RF), `ExtraTreesClassifier` (ET), and `KNeighborsClassifier` (5-NN) (Scikit-learn developers, 2025b;a;c). Each ensemble has five fully grown trees with default parameters (RF bootstraps; ET does not). The neighbor classifier uses five Euclidean neighbors with uniform weights; for 5-NN only, `StandardScaler` is fitted on the training split. Scores are the fractions of tree or neighbor labels voting for each class, so every score lies in $\{0, 1/5, \dots, 1\}$ without post-hoc quantization. Five votes keep exhaustive certification tractable; the study does not establish scalability to dense softmax scores. Score ties are resolved by ascending class index.

Information boundary. For each evaluation batch, the model owner exports the sample size n, class count N, score denominator $M = 5$, tie rule, top-1 hit count b_1, and the separate binary counts

$$b_{i,v,z} = \#\{j : MS_{j,i} = v,\ \mathbf{1}\{X_j = i\} = z\}, \qquad z \in \{0, 1\}.$$

These counts determine the complete empirical one-vs-rest score/label laws. Fitting and export run separately from endpoint optimization; the solver reads only the report.

An exact learned-score coupling problem for every rank. For each normalized integer score tuple $a \in \{0, \dots, M\}^N$ with $\sum_i a_i = M$, and label $y \in [N]$, retain the pair only if $b_{i,a_i,\mathbf{1}\{y=i\}} > 0$ for every i. Call this finite set $\mathcal{K}_b$. Let $t_k(a, y) = \mathbf{1}\{y \in \mathrm{Top}_k(a)\}$, using the specified tie rule. The continuous endpoint program is

$$\begin{aligned} \min/\max_{w \geq 0} \quad & \frac{1}{n} \sum_{(a,y) \in \mathcal{K}_b} w_{a,y} t_k(a, y), \\ \text{subject to} \quad & \sum_{a,y} w_{a,y} = n, \qquad \sum_{a,y} w_{a,y} t_1(a, y) = b_1, \\ & \sum_{(a,y) \in \mathcal{K}_b:\ a_i = v,\ \mathbf{1}\{y=i\} = z} w_{a,y} = b_{i,v,z} \quad \text{for every reported } (i, v, z). \end{aligned} \tag{84}$$

The objective is actual accuracy, not $\mathbb{E}F_k(S)$, and no calibration assumption enters. LP duality supplies certificates of the form

$$d(a, y) = \alpha + \beta t_1(a, y) + \sum_i f_i(a_i, \mathbf{1}\{y = i\}) \ \leq \ t_k(a, y) \quad \text{or} \quad d(a, y) \geq t_k(a, y).$$

Matching primal and dual objectives proves optimality.

Original top-2 statistics over all 150 reports. Table 11 reports mean and sample standard deviation across the ten seeds for observed top-1 and top-2, lower and upper endpoints, and interval width. Figure 7 displays every seed's width, including zeros. All 150 reported intervals contain their corresponding owner-side top-2 value. The study yields 90 positive-width reports and 60 exact-identification reports, with individual widths up to 5.56 percentage points. These counts describe the prespecified study, not the prevalence of ambiguous reports outside it.

Table 10: **Data used by the repeated-holdout protocol.** The same cleaned array and seed-specific split are shared across models. All classes and features are retained. Landsat's two source files are pooled before splitting; these are not official benchmark-split results.

Dataset	Raw rows	Removed	Train	Held out	Features	Classes
Iris	150	1	104	45	4	3
Wine	178	0	124	54	13	3
Digits	1,797	0	1,257	540	64	10
Dry Bean	13,611	68	9,480	4,063	16	7
Landsat	6,435	0	4,504	1,931	36	6

Table 11: **Accuracy and endpoint statistics for all dataset/model cells.** Entries are percentages, mean $\pm$ sample SD over ten seeds; widths are percentage points. A_2 is owner-side accuracy withheld from the solver. L_b, U_b are exact weighted-law endpoints for each report; averaging their endpoints does not define a single pooled report. The last column counts strictly positive exact widths.

Dataset	Model	A_1	A_2	L_b	U_b	Width	> 0
Iris	RF	94.22 ± 5.26	97.56 ± 1.64	97.56 ± 1.64	97.56 ± 1.64	0.00 ± 0.00	0/10
	ET	94.89 ± 2.97	99.11 ± 1.55	99.11 ± 1.55	99.11 ± 1.55	0.00 ± 0.00	0/10
	5-NN	94.44 ± 3.51	99.78 ± 0.70	99.78 ± 0.70	99.78 ± 0.70	0.00 ± 0.00	0/10
Wine	RF	95.37 ± 3.06	99.44 ± 0.89	99.44 ± 0.89	99.44 ± 0.89	0.00 ± 0.00	0/10
	ET	95.37 ± 2.51	99.81 ± 0.59	99.81 ± 0.59	99.81 ± 0.59	0.00 ± 0.00	0/10
	5-NN	94.81 ± 3.12	99.81 ± 0.59	99.81 ± 0.59	99.81 ± 0.59	0.00 ± 0.00	0/10
Digits	RF	91.20 ± 1.42	95.17 ± 1.04	93.52 ± 1.34	97.85 ± 0.73	4.33 ± 0.81	10/10
	ET	92.59 ± 0.86	96.48 ± 0.79	95.19 ± 0.86	98.48 ± 0.53	3.30 ± 0.66	10/10
	5-NN	97.39 ± 0.58	99.09 ± 0.28	98.67 ± 0.30	99.39 ± 0.21	0.72 ± 0.27	10/10
Dry Bean	RF	91.32 ± 0.48	97.11 ± 0.28	96.14 ± 0.36	97.82 ± 0.26	1.68 ± 0.22	10/10
	ET	90.81 ± 0.64	97.05 ± 0.37	96.06 ± 0.43	97.69 ± 0.31	1.63 ± 0.16	10/10
	5-NN	92.25 ± 0.40	97.75 ± 0.19	96.94 ± 0.30	98.17 ± 0.16	1.24 ± 0.22	10/10
Landsat	RF	88.82 ± 0.38	95.78 ± 0.42	94.00 ± 0.47	97.67 ± 0.32	3.67 ± 0.34	10/10
	ET	88.63 ± 0.81	96.06 ± 0.45	94.51 ± 0.53	97.80 ± 0.33	3.29 ± 0.47	10/10
	5-NN	90.79 ± 0.67	97.55 ± 0.18	96.89 ± 0.36	98.23 ± 0.24	1.35 ± 0.30	10/10

Information retained beyond top-1. Top-1 alone gives the elementary actual-top-2 interval $[A_1, 1]$. Table 12 compares its width with the certified report width, computing relative reduction separately for each run before averaging. Runs with perfect top-1 have an undefined relative reduction and are excluded from that ratio only, with the included count reported.

Exactness and interpretation. The report solver and independent rational checker certify both LP endpoints in every run. The admissible domains range from 4 to 13,993 score/label pairs, with 699,138 dual inequalities checked exactly in total. All 300 endpoints have integer witnesses at the original evaluation size.

These intervals concern actual accuracy for score/label laws compatible with the export. The posterior $1/5$ upper theorem does not apply without full vector calibration. Across-seed standard deviations use denominator nine and describe variation in the split/fitting procedure; evaluation sets overlap across seeds, so these are not independent population confidence intervals.

Reproduction. The main commands are:

```
python3 code/classifier_sweep.py run --workers 4
python3 code/check_classifier_sweep.py
python3 code/refit_classifier_sweep.py
```

The sweep checker verifies all reports, witnesses, dual inequalities, source/data hashes, and interval containment. A fresh-process refit of the first seed in every cell exactly reproduces 15 sets of prediction rows. The archive README gives preparation and environment instructions.

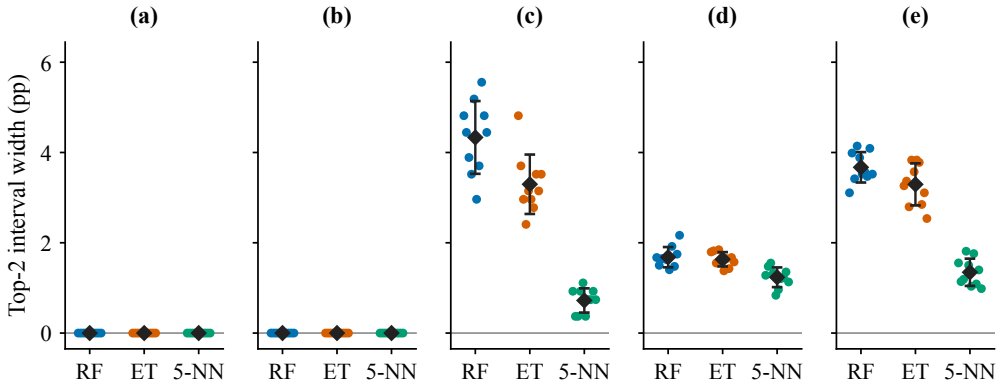


Figure 7: **All seed-level top-2 interval widths.** (a) Iris, (b) Wine, (c) Digits, (d) Dry Bean, and (e) Landsat. Each dot is one report; horizontal offsets separate the ten seeds. RF and ET use five trees; 5-NN uses five neighbors. Widths are in percentage points (pp). Black diamonds and bars show the mean and sample standard deviation. All panels share one vertical scale. Error bars summarize repeated-split variation, not uncertainty about an interval endpoint or a population mean.

Table 12: **Information retained relative to top-1 alone.** The elementary interval $[A_1, 1]$ has width $1 - A_1$. Relative reduction is $100[1 - (U_b - L_b)/(1 - A_1)]$, computed per run then averaged. Perfect-top-1 runs make this ratio undefined; the count of included seeds is displayed. A reduction measures reporting information, not a predictive-performance improvement.

Dataset	Model	Top-1-only width (pp)	Reduction (%)	Included seeds
Iris	RF	5.78 ± 5.26	100.00 ± 0.00	8
	ET	5.11 ± 2.97	100.00 ± 0.00	9
	5-NN	5.56 ± 3.51	100.00 ± 0.00	9
Wine	RF	4.63 ± 3.06	100.00 ± 0.00	9
	ET	4.63 ± 2.51	100.00 ± 0.00	9
	5-NN	5.19 ± 3.12	100.00 ± 0.00	9
Digits	RF	8.80 ± 1.42	50.53 ± 7.00	10
	ET	7.41 ± 0.86	55.38 ± 7.56	10
	5-NN	2.61 ± 0.58	71.49 ± 11.98	10
Dry Bean	RF	8.68 ± 0.48	80.64 ± 2.15	10
	ET	9.19 ± 0.64	82.23 ± 1.02	10
	5-NN	7.75 ± 0.40	84.11 ± 2.30	10
Landsat	RF	11.18 ± 0.38	67.07 ± 3.70	10
	ET	11.37 ± 0.81	71.11 ± 2.74	10
	5-NN	9.21 ± 0.67	85.41 ± 2.88	10

I ALL-RANK COVERAGE AND SHORTLIST-BUDGET STUDY

What is reused and what is new. The data, models, splits, score exports, and owner-side validation records are the 150 reports described in Section H. The extension optimizes the actual label-hit objective in Equation (84) for every $k = 1, \ldots, N$ on these same exports. Iris and Wine have three classes, Landsat six, Dry Bean seven, and Digits ten. This yields 870 report/rank pairs, including 570 with $2 \le k < N$. It is a post-hoc analysis of the frozen reports, not a new training experiment. The original top-2 numerical results are reproduced in all 150 cases.

Exact verification and integer batches. The independent rank checker enumerates normalized integer compositions, retains every feasible score/label pair, and computes its rank by counting strictly larger scores and equal-score classes with lower index. This independently checks the solver's sorted-index rank calculation. For each certificate it reconstructs mass, top-1, every classwise count, and the objective, then checks every dual inequality with rational arithmetic. In total it verifies

Table 13: **Exact actual top-k ambiguity across all nontrivial ranks.** Widths use integer-batch endpoints and are measured in percentage points (pp). Eligible report counts change with k because datasets have different class counts.

k	Reports	Positive width	Mean width (pp)	Max width (pp)
2	150	90	1.414	5.556
3	90	90	2.101	5.926
4	90	90	2.063	5.000
5	90	86	0.964	2.020
6	60	59	0.847	1.600
7	30	28	0.519	1.296
8	30	27	0.481	1.296
9	30	29	0.611	1.667

6,276,148 dual inequalities across all ranks. All supplied top-1 and terminal top-N controls have zero width.

Of 1,740 LP extrema, 1,736 already have integer primal witnesses. Four lower extrema are fractional in hit counts:

Dataset/model	Seed	k	LP hits	Integer hits
Dry Bean/ET	20260917	4	$7861/2$	3931
Landsat/5-NN	20260915	4	$5618/3$	1873
Landsat/5-NN	20260916	4	$3761/2$	1881
Landsat/5-NN	20260920	4	$3787/2$	1894

The exact LP dual lower bound applies to every integer batch. Therefore its ceiling bounds the integer hit count from below. An independently checked integer witness at that ceiling proves the exact corrected endpoint. The largest correction is 0.025894 percentage points; all main-text statistics and budget decisions use these integer endpoints. This argument establishes both extrema over literal batches, but does not make every interior LP value feasible at the same batch size.

Targets, case selection, and summary statistics. Before calculating the budget analysis, we recorded targets $\tau \in \{9/10, 19/20, 99/100\}$, inclusion of all 150 reports, and the illustration rule: the first declared Digits/RF seed, 20260911. The outcome is $b_{\text{certified}} - b_{\text{possible}}$, in candidates. No target, report, or zero-width result is omitted. Table 14 provides every dataset and target. The number with top-1 below target distinguishes decisions already resolved by the exported top-1 statistic. Means average reports equally; repeated splits overlap and supply descriptive variation, not independent sampling replicates or population confidence bounds.

Attaining the two budget extrema. At b_{possible}, we reconstruct the upper-endpoint integer witness; its full top-k curve must first cross the target at that budget. At $b_{\text{certified}} > 1$, we reconstruct the lower-endpoint witness at the preceding rank; its full curve must first cross at $b_{\text{certified}}$. If $b_{\text{certified}} = 1$, any compatible witness suffices. The checker re-forms each score/label row, recomputes its rank using the fixed tie rule, and reconstructs every exported count with Python integers. It verifies endpoint objectives, monotonicity, containment in all rank-wise bounds, and the asserted first crossing. All 900 witness checks pass. No joined owner row is read by this budget computation. The observed budget is taken only from the separately validated owner-accuracy curve and checked for containment between the two extrema.

Reading the illustrative report. For the 540-example Digits/RF report, top-1 is $495/540 = 11/12$. At 95% coverage, the exact minimum-budget extrema are two and five. At 90%, both are one; at 99%, they are eight and ten. The owner-side observed budgets are one, three, and nine at these targets. The curves in Figure 1 are actual feasible endpoint witness curves for the 95% decision, rather than the pointwise upper and lower envelopes passed off as jointly feasible curves.

Limits of the application scenario. The study is about the consequences of an explicitly specified report. It does not assert that existing routing or retrieval systems use this export policy. No constraint

Table 14: **Every dataset and target in the budget study.** *Below* counts reports with top-1 below target; *Ambiguous* counts positive $b_{\text{certified}} - b_{\text{possible}}$. Widths are in candidates, averaged equally over all reports in the row, including zero widths.

Target	Dataset	Reports	Below	Ambiguous	Mean width	Max width
90%	All	150	24	0	0.000	0
90%	Digits	30	1	0	0.000	0
90%	Dry Bean	30	0	0	0.000	0
90%	Iris	30	2	0	0.000	0
90%	Landsat	30	21	0	0.000	0
90%	Wine	30	0	0	0.000	0
95%	All	150	110	29	0.340	3
95%	Digits	30	20	10	0.700	3
95%	Dry Bean	30	30	0	0.000	0
95%	Iris	30	14	0	0.000	0
95%	Landsat	30	30	19	1.000	2
95%	Wine	30	16	0	0.000	0
99%	All	150	143	85	1.187	5
99%	Digits	30	30	29	2.800	5
99%	Dry Bean	30	30	26	1.300	3
99%	Iris	30	26	0	0.000	0
99%	Landsat	30	30	30	1.833	2
99%	Wine	30	27	0	0.000	0

is imposed from a fitted model, feature geometry, or training distribution that the recipient was not given. Five-vote support makes exact enumeration feasible and does not establish scalability to dense scores. The target is shortlist coverage of the recorded batch; a reranker can still choose incorrectly within a covered list. We measure neither second-stage runtime nor its final accuracy.

Reproduction. The code bundle supplies `topk_reporting.py`, the independent `check_topk_report.py`, and the complete certificates under `code/topk_sweep_20260913/`. The budget script `code/shortlist_budget_study.py` reads these frozen artifacts, checks their hashes, applies the four integer corrections, and writes individual decisions, summaries, endpoint-witness curves, and verification records under `code/shortlist_budget_20260913/`. The plotting script reads those checked files directly.

J Target Sweeps and Frozen-Score Report Enrichment

J.1 Exact sweep without additional optimization

The post-hoc target analysis reads the saved integer-batch envelopes for all 150 reports, including the four corrected integer lower endpoints. It first checks the source artifact hashes against the original budget verification record. No model is refitted and no endpoint is reoptimized.

For each report, sort the distinct values in $\{0, 1, L_1, \ldots, L_N, U_1, \ldots, U_N\}$. Between consecutive values $a < b$, both budgets are constant on $(a, b]$. The upper endpoint is included because sufficiency is defined by coverage $\geq \tau$; the lower endpoint belongs to the preceding cell. The checker evaluates both b and the midpoint $(a + b)/2$ using rational arithmetic, verifies the straddling-rank identity in Equation (17), and sums cell lengths times budget gaps. For all 150 reports this integral equals $\sum_{k<N}(U_k - L_k)$ exactly. The finite-cell representation therefore covers all real targets in $(0, 1]$, rather than a grid approximation.

Ninety reports have a positive gap on some target interval, and 60 are identified at every target. The maximum simultaneous count of ambiguous reports is 88; the maximum budget gap is five candidates. No report is ambiguous at or below $11/12$, consistent with all 90% decisions being identified. The mean integrated gap across reports is $32741299777/635497893000 \approx 0.051521$; the mean total length of the ambiguous target set is $14252615653/635497893000 \approx 0.022427$. These are Lebesgue integrals over the coverage scale, not expectations under an estimated distribution of deployment

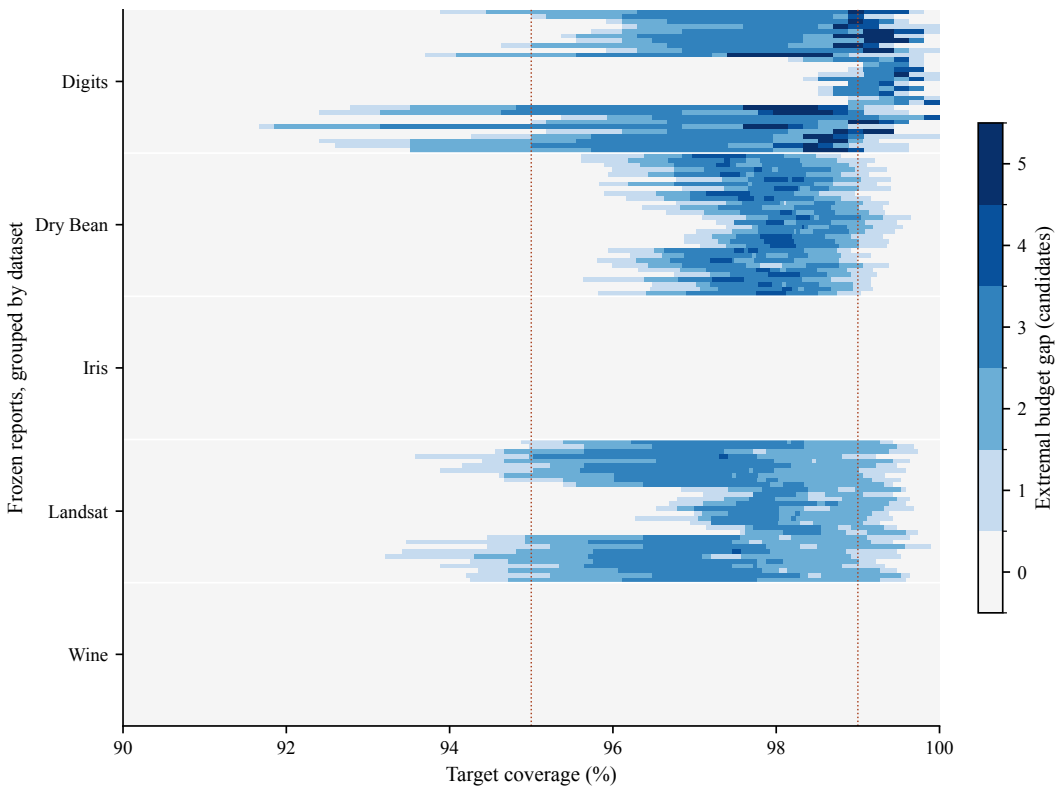


Figure 8: **Where each report leaves a budget unresolved.** Every horizontal row is one of the 150 frozen reports, ordered by dataset, model, and seed. Color gives the exact extremal budget gap on each target cell. All targets below the displayed 90% boundary are identified. Vertical dotted lines mark the original 95% and 99% targets; endpoint inclusion follows $(a, b]$ as described in the text.

targets. For the illustrative report they are $5/27$ and $1/15$, respectively. The original target counts (0, 29, 85 at 90%, 95%, 99%) are reproduced exactly.

The script `code/target_sweep_20260918/analyze.py` generates the exact cell tables, per-report integrals, and cohort sweep underlying the plots. The analysis checks 2,098 cell-endpoint and midpoint evaluations, with input and output hashes saved in `code/target_sweep_20260918/checks.json`.

J.2 Adding information while holding the scoring rule fixed

This focused follow-up uses the existing illustrative report: Digits, random forest, seed 20260911, with 540 evaluation examples. Selection is inherited from the earlier budget analysis. The enrichment ladder, all ten ranks, targets, public grid, and 60-second per-endpoint solver limit were recorded before running this follow-up. The analysis is post-hoc relative to the original experiments and is not a new prespecified benchmark or a prevalence estimate.

Every export supplies $N = 10$, score denominator $M = 5$, batch size 540, and ascending-index tie breaking. The owner computes successively nested summaries from the same frozen rows: (i) top-1 and class frequencies; (ii) additionally all classwise score/label counts; (iii) additionally the multiclass argmax confusion matrix; (iv) additionally the maximum-score histogram; and (v) additionally the true-label rank histogram. No scores, labels, or ranking rules change across levels. Constraints from joint summaries are ordinary linear count equations: for example, confusion entry (y, j) fixes the mass of states with true label y and predicted argmax j.

The exact five-level results are shown in main-text Table 3. Each level fixes the same public score grid; all endpoints are weighted-law extrema.

Table 15: **Cutoff ties on the unchanged evaluated rows.** Counts and percentages separate positive-score and zero-score ties. Digits/RF has 540 rows; rounded RepVGG-A0 has 10,000. The diagnostics characterize the rules and do not decompose interval width causally.

Rule	k	Positive cutoff ties	Zero cutoff ties
Digits/RF, five votes	2	122 (22.59%)	225 (41.67%)
Digits/RF, five votes	5	0 (0%)	534 (98.89%)
Rounded RepVGG-A0, $M = 5$	2	48 (0.48%)	9493 (94.93%)
Rounded RepVGG-A0, $M = 5$	5	0 (0%)	9999 (99.99%)

Avoiding information leakage into the coarse export. The weakest solver receives no empirical coordinate supports, zero counts, confusion entries, or owner rows. Each level uses the same full public domain of $10\binom{14}{9} = 20{,}020$ labeled denominator-five simplex states. A missing feature never becomes a constraint. Once the complete classwise laws are supplied, their zero counts are legitimate equations; they are not silently inherited by the weaker level. Owner export, report-only solving, and checking run as separate commands.

Exact certificates and their scope. The solver supplies a rational primal measure and dual multipliers for both endpoints at all ten ranks and five levels: 100 endpoints. A verifier enumerates all score rows independently using stars and bars, computes ranks by pairwise comparison, reconstructs every reported count, and checks all 2,002,000 dual inequalities with denominators cleared. Every primal and dual objective agrees exactly. It also checks nested intervals, containment of the observed owner curve, and agreement of the classwise level with the previously certified report endpoints. The solver's supplied witnesses are integer at the weakest level and fractional at the four richer levels; this table therefore claims weighted-law extrema. The original integer classwise witnesses remain available, and the actual owner batch attains the rank-histogram curve. We do not infer integer attainability of enriched extrema from integral objective values alone.

At 95%, adding the confusion matrix reduces the certified budget from five to four without identifying it. The maximum-score histogram further reduces integrated coverage uncertainty but leaves the displayed budget pairs unchanged. These are case-specific findings; the experiment does not prove either general sufficiency or general insufficiency of those joint summaries. Saved reports, certificates, protocol, and checker are in `code/report_enrichment_20260918/`. Reproduce with `study.py export`, `study.py solve`, and `study.py check`.

J.3 Positive-score and zero-score cutoff ties

The owner-side diagnostic records $s_{(k)} = s_{(k+1)} > 0$ separately from $s_{(k)} = s_{(k+1)} = 0$. These events concern the shortlist boundary, not arbitrary ties elsewhere in a row. With denominator five, no row can have more than five positive coordinates; in ten dimensions every cutoff from six through nine is necessarily inside the zero-score block.

The owner diagnostic script `code/target_sweep_20260918/ties.py` reconstructs top-1 and every classwise count from the saved rows, verifies the neural totals against the original cutoff-tie audit, and saves input hashes and all nonterminal cutoffs. It is separate from both endpoint solvers. It does not assert that these frequencies hold for the compatible extremal laws or for the original unrounded neural scores.

J.4 Sensitivity to alternative fixed tie rules

This limited post-hoc analysis keeps the 540 score vectors and true labels of the illustrative Digits/RF batch fixed. Before optimization, we record three deterministic priorities: ascending $(1, \ldots, 10)$, descending $(10, \ldots, 1)$, and cyclic $(6, 7, 8, 9, 10, 1, 2, 3, 4, 5)$, using one-based class indices. These priorities are independent of the score vector and true label. For each priority we export its actual top-1 count, recompute all classwise score/label counts, and use that same priority for every endpoint objective and owner audit. The classwise counts remain unchanged; top-1 and coverage need not. Thus this is a comparison of consistently recomputed reports, not a change of objective under the old report.

Table 16: **One frozen batch under three fixed tie priorities.** Budget pairs are weighted-law extrema at the indicated targets. The integral is $\sum_{k<10}(u_k - \ell_k)$. All 60 endpoint certificates pass exact primal/dual checks; integer-batch attainment is not asserted here.

Priority	Top-1	90% budgets	95% budgets	99% budgets	Integral
Ascending	$11/12$	$(1, 1)$	$(2, 5)$	$(8, 10)$	$5/27$
Descending	$241/270$	$(2, 2)$	$(2, 4)$	$(5, 10)$	$26/135$
Cyclic	$97/108$	$(2, 2)$	$(2, 4)$	$(5, 10)$	$17/90$

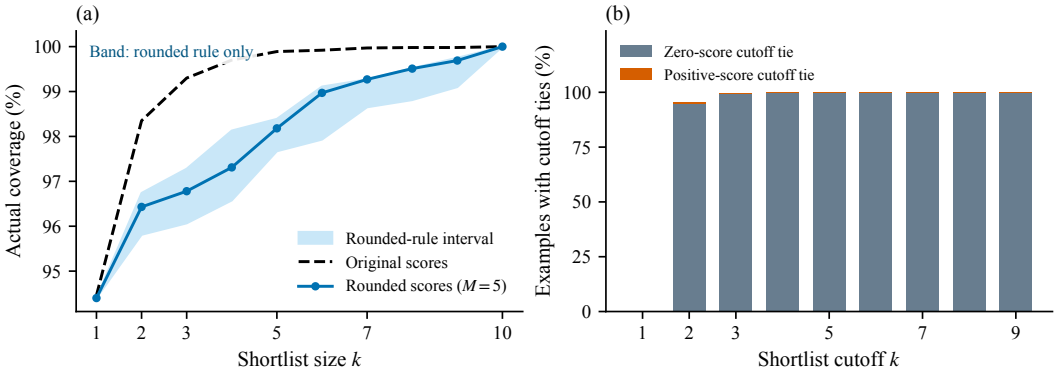


Figure 9: **Score coarsening changes the decision being certified.** Left: original and rounded RepVGG-A0 coverage; the band certifies only the rounded rule ($M = 5$). Right: cutoff ties under that rule, separating equal positive scores from zero scores. The large zero-score blocks make the fixed tie rule consequential. Full protocols and costs: Section K; tie counts: Table 15.

The 95% owner budgets are 3, 2, and 3; their 99% budgets are 9, 7, and 8. Budget ambiguity at 95% persists for these three rules in this case, while both the certified range and the owner decision can change. The 90% decision remains identified but its value changes with top-1. These findings do not establish tie robustness for the 150-report cohort or attribute uncertainty to ties alone.

The solver reads only each exported report and uses the full public domain of 20,020 labeled denominator-five simplex states. A separate checker uses stars-and-bars enumeration and pairwise comparisons for ranks, checks every primal constraint and all 1,201,200 dual inequalities exactly, and recovers the previously certified ascending-order endpoints. The supplied witnesses are fractional, so the analysis reports weighted-law extrema. Scripts, protocol, report inputs, certificates, and checks are in `code/tie_sensitivity_20260918/`; run `study.py export`, `study.py solve`, and `study.py check` as separate commands.

K Neural Classifier, Score Resolution, and Computational Scaling

A frozen neural classifier on a separate image benchmark. We evaluate a public CIFAR-10 RepVGG-A0 checkpoint (Ding et al., 2021; Chen, 2026) on the complete official 10,000-image test batch (Krizhevsky, 2009). The checkpoint, architecture revision, test archive, and image normalization are fixed before inference and preserved with source and data hashes. We perform no retraining, fine-tuning, temperature fitting, or checkpoint selection. This is one externally trained neural classifier, not a new set of independent training replicates. It is a separate cohort from the 150 classical-model reports in Section H. Images retain their native 32×32 resolution, are divided by 255, and use the provider's channel means $(0.4914, 0.4822, 0.4465)$ and standard deviations $(0.2023, 0.1994, 0.2010)$. Inference uses evaluation mode and unmodified pretrained weights; saved logits permit all later analyses without rerunning the network.

Precision changes the evaluated score rule. Let $p = \mathrm{softmax}(z)$ at temperature one. For denominator M, we first take $v_i = \lfloor Mp_i \rfloor$, then add one to the $M - \sum_i v_i$ coordinates with largest

fractional remainders. Equal remainders and equal rounded scores use ascending class index. The evaluated finite-support classifier has scores v/M; both its exported top-1 and its top-k objectives use this same rounded rule. The owner separately records actual label-hit accuracies under the original scores. We also measure the fraction of examples whose top-k *set* changes and the fraction with a tie across the shortlist cutoff. Thus an interval certified for rounded scores is never presented as an interval for the original softmax classifier. This is output-score rounding, not neural-weight or activation quantization. The different M settings need not define nested reports or nested feasible sets, so monotone narrowing is not assumed.

Design recorded before inference. The primary experiment uses all ten classes, $M = 5$, all 10,000 test examples, and every $k = 1, \dots, 10$. The resolution grid crosses $M \in \{2, 5, 10, 20, 50, 100\}$ with $N \in \{3, 5, 10\}$. For the smaller-class controls we retain test labels $0, \dots, N-1$ and softmax-normalize the corresponding logits, yielding $1,000N$ examples. These are controlled restricted-label tasks; they do not measure the original ten-class classifier's accuracy. Every cell evaluates $k = 2$; the ten-class cells also evaluate $k = 5$. A batch-size study fixes $N = 10, M = 5$ and uses $B \in \{100, 500, 1000, 5000, 10000\}$. Within each class, a PCG64 permutation with seed 20260913 supplies nested subsets of $B/10$ examples. The full-batch cell is shared with the resolution grid. In total the design contains 22 distinct reports. Three fresh-process timing runs are specified for the $B = 1000$ case; other timings are single-run, hardware-specific measurements. Overlapping subsets and timing repetitions do not constitute independent model fits.

Counting the full domain before allocating it. For label y, let $A_i^{(y)}$ be the integer scores with positive reported count in the event $\mathbf{1}\{y = i\}$. The exact number of admissible labeled states is

$$D = \sum_{y=0}^{N-1} [x^M] \prod_{i=0}^{N-1} \left(\sum_{v \in A_i^{(y)}} x^v \right) \le N \binom{M+N-1}{N-1}.$$

A dynamic program multiplies these support polynomials, truncating degrees above M, with Python integers. This gives an exact count even when explicit enumeration is disallowed by the resource budget. It does not solve the endpoint LP. The predeclared limits are 10^6 labeled states, 180 seconds per LP, 600 seconds per report process, and 8 GiB process RSS. A domain above the state limit is marked *capped before solving*; it is neither a timed-out LP nor an identified report. Every such case remains in the result tables, without an imputed endpoint or zero width. Numerical LP success is also insufficient: a failed exact rational reconstruction is recorded as a certification failure, with its actual time and memory consumption. This is a limitation of the implementation, not a proof of computational hardness or a statement about the report's true interval.

Solver isolation and verification. The endpoint process reads only the report containing all classwise score/label counts and rounded top-1. It enumerates every supported normalized score/label tuple, constructs the coupling LP in Equation (84), and optimizes the actual hit objective. Array-based sparse-matrix construction reduces implementation overhead without restricting the feasible domain. The unchanged independent checker enumerates integer compositions and checks every rational dual inequality, as well as every primal marginal, mass, top-1 count, and endpoint objective. Its enumeration must also match the dynamic-program count. An owner audit independently reconstructs rounding, labels, indices, report counts, observed accuracies, and interval containment. We report weighted-law endpoints and the integer-primal flag for each endpoint. An integral hit count alone is not evidence of an attainable integer batch; we claim literal-batch attainment only when all primal counts are integers.

Timing scope. Each case starts a fresh numerical-solver process within a four-CPU Slurm allocation, with OpenMP and BLAS thread environment variables set to one. End-to-end wall time includes interpreter startup, exact domain counting, enumeration, sparse-matrix construction, independent checker preparation, all requested LPs, rational certificate construction, exact verification, and output writing. Inference and owner-side export are excluded and reported separately. Phase timings distinguish construction from solving and verification for completed cases; unsuccessful attempts retain their total wall time and sampled memory. Peak RSS uses the process high-water mark for completed runs and 100-ms sampling for unsuccessful attempts. Counting-only cases can finish before

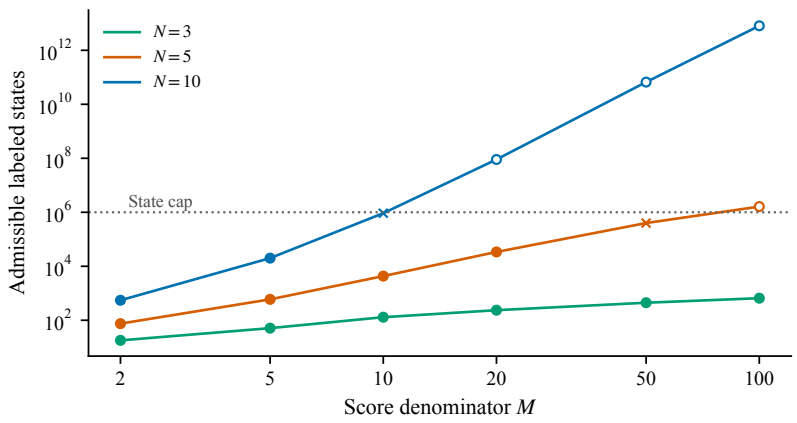


Figure 10: **State counts for exhaustive certification.** Open circles exceed the declared 10^6-state cap; crosses mark unsuccessful certification attempts. Neither receives an interval. The $N = 3, 5$ settings are restricted-label computational controls, not original ten-class accuracy comparisons. Complete costs and failures appear below.

reliable memory sampling, so their RSS is omitted. CPU model, software versions, and repeated timing results are saved with the artifacts. Counting-only capped cases are explicitly distinguished from completed certification when comparing costs.

Observed completion and cost. Sixteen reports complete certification, yielding 60 distinct endpoints and 600,382 independently checked dual inequalities; 26 endpoint witnesses have integer counts. Four reports exceed the state cap. At $N = 10, M = 10$, the top-2 upper LP hits its 180-second limit: the unsuccessful report attempt takes 234.51 seconds and peaks at 2,647.8 MiB. At $N = 5, M = 50$, rational primal reconstruction fails the exact mass/top-1 equations after 146.33 seconds and 851.4 MiB. Neither case supplies a certified interval. These are reported outcomes of the fixed implementation and budget, not evidence that the associated mathematical problems cannot be solved.

The common $k = 2, 5$ workload at $N = 10, M = 5, B = 10000$ takes 3.52 seconds and 129.6 MiB; the primary all-ten-rank run takes 21.68 seconds. The three $B = 1000$ timing repetitions range from 1.508 to 1.512 seconds and reproduce the same endpoint values. Across $B = 100, 500, 1000, 5000, 10000$, domain sizes are 19, 1,283, 7,898, 20,020, and 20,020. Thus sample size affects this solver through which score supports appear and through the constraint counts; it does not add one LP variable per observed example. These timings use an Intel Xeon Platinum 8358P at 2.60 GHz, Python 3.11.15, NumPy 2.2.6, SciPy 1.17.1, and the same four-CPU allocation. The separate pretrained inference/export job completes in 92 seconds, using PyTorch 2.8.0 on CPU and a network with 7,840,874 parameters.

Reproduction and scope. The directory `code/neural_resolution_20260913/` contains the pre-inference protocol, public-input hashes and provider code, saved owner logits and exports, all 22 reports, solver outputs including capped cases, exact certificates, resource records, and scripts that regenerate the audit tables. `neural_study.py` separates inference/export from report-only solving; `analyze_study.py` reconstructs the exported counts and rechecks every completed certificate. The study tests a finite reporting boundary for a neural classifier and maps the practical limits of this exhaustive implementation. It does not establish efficient certification for dense, unrounded scores or large-label deployment, nor does it infer population coverage from this fixed evaluation batch.

Table 17: RepVGG-A0 on the full ten-class test batch, $M = 5$. Percentages are actual label-hit rates. Bounds concern the rounded classifier's weighted report laws. I/F describe integer/fractional counts in the *supplied optimizer witness* at each endpoint, not the possibility of integer attainment. The observed rounded batch is an integer witness for both fixed controls ($k = 1, 10$), even where the supplied optimizer has F/F. Displayed bounds are rounded outward.

k	Original	Rounded	Lower	Upper	Width (pp)	Witness L/U
1	94.46	94.40	94.40	94.40	0.000	F/F
2	98.35	96.43	95.81	96.74	0.930	F/F
3	99.30	96.78	96.06	97.28	1.215	F/F
4	99.70	97.31	96.57	98.13	1.555	F/F
5	99.89	98.18	97.67	98.39	0.720	I/F
6	99.92	98.97	97.93	99.11	1.180	F/I
7	99.97	99.27	98.65	99.27	0.620	F/F
8	99.98	99.51	98.81	99.52	0.710	F/F
9	99.98	99.69	99.10	99.75	0.650	I/F
10	100.00	100.00	100.00	100.00	0.000	F/F

Table 18: Complete resolution grid. D is the exact admissible labeled-state count; W_2 is in percentage points. Time includes fresh-process setup, solving, and exact verification for $k = 2$ ($N = 3, 5$) or $k = 2, 5$ ($N = 10$). Capped cells report no interval, solve time, or solver memory; their fast domain-counting costs remain in the machine-readable table. Check failed denotes rejection by the rational verifier; LP limit denotes a timeout.

N	M	D	W_2	Time (s)	RSS (MiB)	Status
3	2	18	0.100	0.61	66.5	Certified
3	5	51	0.000	0.60	66.8	Certified
3	10	130	0.033	0.61	67.1	Certified
3	20	236	0.000	0.71	67.3	Certified
3	50	446	0.067	0.61	68.4	Certified
3	100	657	0.000	0.71	69.5	Certified
5	2	75	0.180	0.60	67.1	Certified
5	5	595	0.500	0.71	68.4	Certified
5	10	4,325	0.680	0.81	77.5	Certified
5	20	33,918	0.900	8.03	155.9	Certified
5	50	398,749	–	146.33	851.4	Check failed
5	100	1,626,181	–	–	–	Capped
10	2	550	0.560	0.71	69.2	Certified
10	5	20,020	0.930	3.52	129.6	Certified
10	10	923,375	–	234.51	2647.8	LP limit
10	20	90,318,123	–	–	–	Capped
10	50	66,422,884,687	–	–	–	Capped
10	100	8,079,838,613,038	–	–	–	Capped

Table 19: Batch-size scaling for the same rounded classifier ($N = 10, M = 5$). Every timing solves both endpoints at both $k = 2$ and $k = 5$, including a separate two-rank run for $B = 10,000$. Widths are in percentage points.

B	D	W_2	W_5	Time (s)	RSS (MiB)
100	19	1.000	1.000	0.71	66.9
500	1,283	1.000	1.400	0.91	71.6
1,000	7,898	1.100	1.000	1.51	93.4
5,000	20,020	0.940	0.720	3.52	129.7
10,000	20,020	0.930	0.720	3.52	129.6

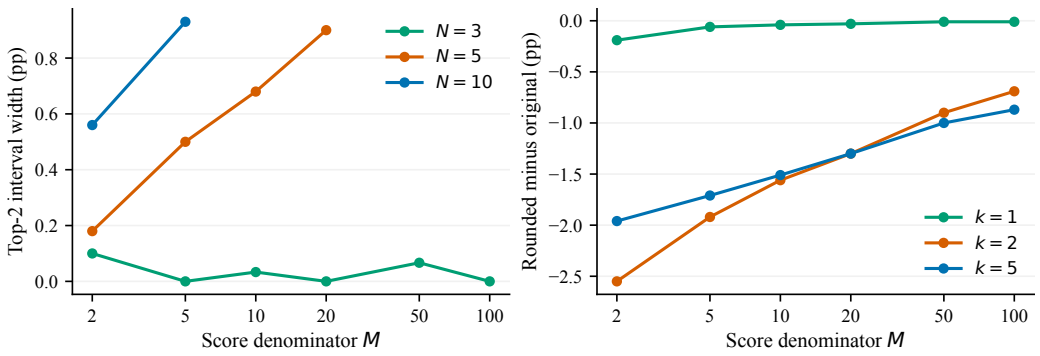


Figure 11: Left: top-2 widths at every certified resolution-grid cell; capped and unsuccessful cases have no reported width. Right: accuracy changes caused by score rounding on the original ten-class task. The owner can compute these changes at every M, including values at which the report-only solver is capped. Preserving top-1 accuracy does not imply preserving larger lists.

L REPRODUCIBILITY

The computational supplement contains exact rational-arithmetic checkers and a standard-library Python runner. From its extracted root, first run `python3 -B verify_package.py`; some later analyses regenerate outputs. Its README supplies complete commands and dependencies. Main checkers in `code/` include:

All 150 classifier reports	`check_classifier_sweep.py`
Individual classifier reports	`check_classifier_report.py`
Fixed-dimension potentials	`check_fixed_n_upper.py`
Report endpoints & relaxations	`check_report_endpoints.py`
Finite-grid certificates	`check_finite_grid_lp.py`
Diagnostic-law equivalence	`check_diagnostic_reports.py`

The certificate verifiers use exact standard-library arithmetic. `check_topk_theory.py` additionally uses NumPy and SciPy for small API control problems. Optional rediscovery scripts (`report_endpoints.py`, `finite_grid_lp.py`) require NumPy and SciPy; they refuse to save results unless exact verification succeeds. Table builders read only verified certificate/checker hash pairs.

The runner additionally covers the potential and size-biased algebra, Bernstein inversion, a bounded row sweep, the five-row and finite-N constructions, anchor-lift identities, fixed-m count equations, and the common-prior multigraph realization. These computations provide regression and falsification evidence; continuum claims are proved symbolically.

BLOCK IV: OPTIONAL EXTENSIONS

M INFORMATION SUMMARIES AND POSTERIOR PROFILES

Representation quality and the evaluator. Information-bottleneck methods balance predictive information and compression (Alemi et al., 2017); predictive $\mathcal{V}$-information and decodable information specify a predictor family (Xu et al., 2020; Dubois et al., 2020). Studies of information objectives, probes, and downstream embeddings show that evaluation depends on the chosen task and predictor (Tschannen et al., 2020; Lovering et al., 2021; Antverg & Belinkov, 2022; Darrin et al., 2024; Radford et al., 2021). Loss-relative comparisons and representation learning for future tasks make this dependence explicit (Duchi et al., 2018; Zhao et al., 2022; Uzan & Weinberger, 2024). Our label task and decoder remain fixed; the restriction is the information available to the report's recipient.

Posterior laws and information summaries. Blackwell equivalence preserves all Bayes decision values (Blackwell, 1953), and quantitative-information-flow hyper-distributions retain the full posterior law (McIver et al., 2014). Coordinate marginals retain less information, so matching them and top-1 need not imply Blackwell equivalence. Entropy-based list-error bounds use Arimoto/up-arrow conditional entropy (Sason & Verdú, 2018) or generalized Fano inequalities (Sakai, 2020). Our appendix instead studies the complete positive-order Petz/down-arrow profile, using Tomamichel et al. (2014)'s conventions. Its diameter result does not assert an Arimoto analogue.

From the Rényi profile to the pooled measure. For $\alpha > 0$, let

$$M_\alpha(\mu) = \int x^\alpha \, \nu_\mu(\mathrm{d}x), \qquad H_\alpha^\downarrow(\mu) = \frac{\log M_\alpha(\mu)}{1-\alpha} \quad (\alpha \neq 1).$$

At $\alpha = 1$, set $H_1^\downarrow(\mu) = -\int x \log x \, \nu_\mu(\mathrm{d}x)$, using natural logarithms. This is the continuous Shannon extension; the dominated-limit argument is given below. The moments also satisfy

$$\int x^j \, \lambda_\mu(\mathrm{d}x) = M_{j+1}(\mu), \qquad j \geq 0. \tag{85}$$

The complete positive-order Rényi curve thus reparameterizes the pooled coordinate moments. Its integer orders identify the probability law λ_μ on $[0,1]$. A known dimension also recovers the mass at zero, which is invisible to the size-biased law.

Lemma M.1 (What the complete profile identifies)**.** *For posterior laws in arbitrary finite dimensions, the following are equivalent: equality of $H_n^\downarrow$ at every integer order $n \geq 2$, equality of their true-class-confidence laws λ_μ, and equality of their indexed curves $\alpha \mapsto H_\alpha^\downarrow$ at all positive orders. If the laws lie on a common known Δ_N, these conditions are also equivalent to equality of their zero-inclusive pooled coordinate measures ν_μ.*

Proof. By Equation (85), the integer orders $n \geq 2$, together with the automatic zeroth moment of the probability λ_μ, identify λ_μ by polynomial density on $[0,1]$ (Hausdorff, 1923). Conversely, $M_\alpha = \int_{(0,1]} x^{\alpha-1} \, \mathrm{d}\lambda_\mu$, so λ_μ gives the full positive-order curve, including the Shannon extension. It determines ν_μ on $(0,1]$ by division by x; a known common N then determines the remaining mass at zero. Without a known dimension, only added zero atoms are invisible. □

Corollary M.2 (Complete-profile diameter and minimax error)**.** *Replace the labeled marginal summary by*

$$\mathsf{S}_{\mathrm{prof}}(\mu) = \left(N, p_\mu, (H_\alpha^\downarrow(\mu))_{\alpha>0}, R_1(\mu)\right), \qquad p_\mu = \int q \, \mathrm{d}\mu.$$

Over posterior laws in all finite label dimensions, the worst-case top-2 diameter is $1/5$, and the minimax expected absolute error of possibly randomized summary-only rules is $1/10$. The midpoint $\frac{2}{5}R_1 + \int \phi \, \mathrm{d}\nu_\mu + \frac{1}{10}$ attains the upper bound.

Proof. At known N, lemma M.1 recovers ν_μ, so the residual proof and midpoint apply unchanged. The finite symmetrized pairs in theorem B.4 match this entire summary, including the uniform prior, and approach a $1/5$ gap. The same two-point triangle inequality as in theorem B.5 proves the minimax lower bound. This concerns Petz/down-arrow entropy with dimension, prior, and top-1 supplied; it asserts no profile-alone or Arimoto theorem. □

Size-biased profile dictionary. Equation (19) gives

$$\lambda_\mu([0,1]) = \int_{\Delta_N} \sum_i q_i \, \mu(\mathrm{d}q) = 1,$$

and, for every integer $r \geq 1$,

$$\int x^r \, \lambda_\mu(\mathrm{d}x) = \int x^{r+1} \, \nu_\mu(\mathrm{d}x) = M_{r+1}(\mu).$$

For $n \geq 2$, natural logarithms give $M_n(\mu) = \exp(-(n-1)H_n^{\downarrow}(\mu))$. Thus equality of all integer profiles gives equality of every positive integer moment of the two size-biased probability laws. Their zeroth moments are one, so polynomial density on $[0,1]$ identifies the measures.

Conversely, equality of the size-biased laws gives, for every $\alpha > 0$,

$$M_\alpha(\mu) = \int_{(0,1]} x^{\alpha-1} \lambda_\mu(\mathrm{d}x).$$

These integrals are finite in every finite dimension: for $0 < \alpha < 1$, $\sum_i q_i^\alpha \leq N^{1-\alpha}$, while for $\alpha \geq 1$, $\sum_i q_i^\alpha \leq 1$. For the Shannon endpoint, when $\alpha \in [1/2, 3/2]$ and $x \in [0,1]$,

$$|x^\alpha \log x| \leq x^{1/2} |\log x| \leq \frac{2}{e},$$

with the value at zero defined by continuity. Since $\nu_\mu([0,1]) = N < \infty$, dominated differentiation and l'Hôpital's rule give

$$\lim_{\alpha \to 1} \frac{\log M_\alpha(\mu)}{1-\alpha} = -\int_{(0,1]} \log x \, \lambda_\mu(\mathrm{d}x).$$

This integral is finite also from $-\sum_i q_i \log q_i \leq \log N$. Finally, multiplication by x is invertible away from zero:

$$\nu_\mu(A) = \int_A x^{-1} \lambda_\mu(\mathrm{d}x), \qquad A \subseteq (0,1].$$

At known N, the remaining mass is $\nu_\mu(\{0\}) = N - \int_{(0,1]} x^{-1} \lambda_\mu(\mathrm{d}x)$. Thus only zero-atom multiplicity is lost without a known dimension.

N Equal-Mass Codebook Budgets

The global theorem permits finite representation alphabets of arbitrary size. We now fix a balanced codebook with exactly m equiprobable outputs and ask when its worst-case gap reaches the global $1/5$ supremum under matched complete profiles and top-1. This is a restriction on output counts and masses; it does not describe typical continuous representations.

An m-row spectrum array $\mathbf{q} = (q^1, \ldots, q^m) \in \Delta_N^m$ induces the empirical posterior law $m^{-1} \sum_r \delta_{q^r}$ and intensity

$$\nu_{\mathbf{q}} = \frac{1}{m} \sum_{r=1}^{m} \sum_{i=1}^{N} \delta_{q_i^r}. \tag{86}$$

Thus equal intensity is a common flattened inventory. Let Γ_m^{sp} be the supremum of $|\Delta R_2|$, over finite N, between equal-inventory arrays in a common Δ_N with equal R_1; common zero padding is allowed. Define Γ_m^{ch} analogously for channel pairs with a common prior, equal complete profiles and R_1, and exactly m labeled outputs of mass $1/m$ per side. Outputs have positive probability and repeated rows are allowed; N, profile, and prior may vary, but m and equal masses are fixed.

Lemma N.1 (Equal-codeword common-prior realization)**.** *For every $m \geq 1$,*

$$\Gamma_m^{\mathrm{ch}} = \Gamma_m^{\mathrm{sp}}. \tag{87}$$

Every equal-inventory pair can be relabeled within rows and realized under one common prior with exactly m equiprobable outputs per channel.

Occurrence matching and perfect-matching factorization assign common labels and column sums; Section P proves both directions. The relabeling may change pre-existing class semantics, and the resulting common prior need not equal a prescribed prior.

Write Γ_m for the common value. In the limiting argument below, *dust* is mass carried by coordinates that vanish individually as the dimension grows; Section O.1 defines the resulting limiting spectra precisely.

Theorem N.2 (Exact equal-mass codebook supremum)**.** *For every integer $m \geq 1$,*

$$\boxed{\Gamma_m = \frac{1}{5} \quad \Longleftrightarrow \quad 5 \mid m.} \tag{88}$$

This is a supremum over finite dimensions, not a finite-attainment claim. Every rowwise dusty accumulation point of an oriented admissible spectral sequence whose signed R_2 gap tends to $1/5$ has the unique count ray in Equation (121); this is subsequential rigidity, not stability or whole-sequence convergence.

Proof idea. Residual endpoint forcing, dusty compactness, and positive-inventory passage give five row types whose count equations force $5 \mid m$; see Section O. Conversely, repeat the five-row witness $m/5$ times and apply lemma N.1. □

The exact-codebook supremum is therefore nonmonotone in m. When $5 \nmid m$, its deficit from $1/5$ is strictly positive; quantifying that deficit remains open.

O Proof of the Equal-Mass Codebook Theorem

The proof proceeds through compactness, contact-set classification, passage of the inventory constraints to the limit, and a final counting argument. These steps establish necessity in Theorem N.2; the repeated five-row construction proves sufficiency. Throughout,

$$\phi(t) = \begin{cases} \frac{8}{5}t^2, & 0 \leq t \leq \frac{1}{4}, \\ \frac{4}{5}t - \frac{1}{10}, & \frac{1}{4} \leq t \leq \frac{1}{2}, \\ \frac{2}{5}t + \frac{1}{10}, & \frac{1}{2} \leq t \leq \frac{3}{4}, \\ -\frac{4}{5} + \frac{14}{5}t - \frac{8}{5}t^2, & \frac{3}{4} \leq t \leq 1. \end{cases} \tag{89}$$

O.1 Dusty compactness and endpoint forcing

Fix m and an oriented sequence satisfying the hypothesis of theorem N.2. Sort every row decreasingly and append infinitely many zeros. For every row label r and rank i, the coordinate lies in $[0, 1]$. Since there are only $2m$ rows and countably many ranks, repeated Bolzano–Weierstrass extraction followed by a diagonal argument gives a common subsequence such that

$$q^{\mathsf{E}}_{n,r,i} \longrightarrow x^{\mathsf{E}}_{r,i} \qquad (\mathsf{E} \in \{\mathsf{H}, \mathsf{L}\},\ r \in [m],\ i \geq 1). \tag{90}$$

The limits remain nonincreasing. For each finite K,

$$\sum_{i=1}^{K} x^{\mathsf{E}}_{r,i} = \lim_{n\to\infty} \sum_{i=1}^{K} q^{\mathsf{E}}_{n,r,i} \leq 1.$$

Letting $K \to \infty$ gives $\sum_i x^{\mathsf{E}}_{r,i} \leq 1$, so

$$d^{\mathsf{E}}_r := 1 - \sum_i x^{\mathsf{E}}_{r,i} \geq 0 \tag{91}$$

defines a dusty spectrum. This argument permits arbitrary numbers and scales of vanishing atoms. For $\bar{q} = ((x_i)_{i\geq 1}, d)$, with missing coordinates set to zero, define its residual explicitly by

$$\bar{h}(\bar{q}) := x_1 + x_2 - \frac{2}{5}x_1 - \sum_{i\geq 1} \phi(x_i). \tag{92}$$

For a sorted unit-mass row,

$$iq_i \le \sum_{j=1}^{i} q_j \le 1, \qquad \text{hence} \qquad q_i \le \frac{1}{i}. \tag{93}$$

When $K \ge 4$, every $i > K$ lies in the first branch of (89), and therefore

$$\sum_{i>K} \phi(q_i) = \frac{8}{5} \sum_{i>K} q_i^2 \le \frac{8}{5(K+1)} \sum_{i>K} q_i \le \frac{8}{5(K+1)}. \tag{94}$$

The same estimate holds for the limiting sequence (x_i), since $x_i \le 1/i$ and $\sum_i x_i \le 1$. For fixed K, continuity of ϕ permits passage to the limit in the first K terms. Applying (94) to both tails and then letting $K \to \infty$ proves

$$\sum_i \phi(q^{\mathsf{E}}_{n,r,i}) \longrightarrow \sum_i \phi(x^{\mathsf{E}}_{r,i}). \tag{95}$$

The first two order statistics converge directly, so

$$h(q^{\mathsf{E}}_{n,r}) \longrightarrow \bar{h}\big((x^{\mathsf{E}}_{r,i})_{i\ge 1}, d^{\mathsf{E}}_r\big). \tag{96}$$

At each n, put $\Delta_n = R_2(\mu_{\mathbf{q}^{\mathsf{H}}_n}) - R_2(\mu_{\mathbf{q}^{\mathsf{L}}_n})$. Common inventory and equal R_1 give the residual cancellation identity

$$\frac{m}{5} - m\Delta_n = \sum_{r=1}^{m} \left(\frac{1}{5} - h(q^{\mathsf{H}}_{n,r})\right) + \sum_{r=1}^{m} h(q^{\mathsf{L}}_{n,r}). \tag{97}$$

Thus (97) is a finite sum of nonnegative terms tending to zero. In particular, for every row label r,

$$h(q^{\mathsf{H}}_{n,r}) \longrightarrow \frac{1}{5}, \qquad h(q^{\mathsf{L}}_{n,r}) \longrightarrow 0. \tag{98}$$

Combining (96) with (98) places every high limit row on the upper contact set and every low limit row on the lower contact set. The conclusion holds for every rowwise dusty convergent subsequence, including but not limited to the diagonal subsequence constructed above.

Finally, equality of R_1 survives the limit. Indeed, at every finite stage,

$$\sum_{r=1}^{m} q^{\mathsf{H}}_{n,r,1} = \sum_{r=1}^{m} q^{\mathsf{L}}_{n,r,1},$$

and (90) permits passage through this finite sum:

$$\sum_{r=1}^{m} x^{\mathsf{H}}_{r,1} = \sum_{r=1}^{m} x^{\mathsf{L}}_{r,1}. \tag{99}$$

O.2 Complete classification of the two contact sets

Set

$$g(t) := \phi(t) \quad (0 \le t \le 1/2), \qquad \eta(t) := t - g(t). \tag{100}$$

On the two branches,

$$\eta'(t) = \begin{cases} 1 - \dfrac{16}{5}t, & 0 < t < 1/4, \\ \dfrac{1}{5}, & 1/4 < t < 1/2. \end{cases}$$

Thus η is strictly increasing on $[0, 1/2]$.

The upper contact set. Let $x = x_1$ and $y = x_2$. Every nonmaximum atom is at most $1/2$, and

$$\bar{h} = \frac{3}{5}x + y - \phi(x) - g(y) - \sum_{i\geq 3} g(x_i). \tag{101}$$

If $x \leq 1/2$, then $y \leq x$, so

$$\begin{aligned} \bar{h} &\leq \frac{3}{5}x - g(x) + \eta(y) \\ &\leq \frac{3}{5}x - g(x) + \eta(x) \\ &= \begin{cases} \frac{8}{5}x - \frac{16}{5}x^2, & 0 \leq x \leq 1/4, \\ \frac{1}{5}, & 1/4 \leq x \leq 1/2. \end{cases} \end{aligned} \tag{102}$$

The first branch reaches $1/5$ only at $x = 1/4$. Equality at $1/5$ in (102) requires

$$x \in [1/4, 1/2], \qquad y = x, \qquad x_i = 0 \ \ (i \geq 3).$$

The missing mass is then $d = 1 - 2x$.

If $x \geq 1/2$, then $y \leq 1 - x \leq 1/2$. The identity

$$\phi(x) + g(1-x) = \frac{4}{5} - \frac{2}{5}x \qquad (1/2 \leq x \leq 1) \tag{103}$$

and strict monotonicity of η yield

$$\begin{aligned} \bar{h} &\leq \frac{3}{5}x - \phi(x) + \eta(y) \\ &\leq \frac{3}{5}x - \phi(x) + \eta(1-x) = \frac{1}{5}. \end{aligned} \tag{104}$$

Equality requires $y = 1 - x$ and $x_i = 0$ for $i \geq 3$. Since $x + y = 1$, it also requires $d = 0$. Direct substitution proves the converse in both cases, establishing

$$\begin{aligned} \{\bar{q} : \bar{h}(\bar{q}) = 1/5\} = \{((s, 1-s), 0) : 1/2 \leq s \leq 1\} \\ \cup \{((s, s), 1 - 2s) : 1/4 \leq s \leq 1/2\}. \end{aligned} \tag{105}$$

The lower contact set when $y > 0$. Define

$$\rho(t) := \frac{g(t)}{t} = \begin{cases} \frac{8}{5}t, & 0 < t \leq 1/4, \\ \frac{4}{5} - \frac{1}{10t}, & 1/4 \leq t \leq 1/2. \end{cases} \tag{106}$$

The function ρ is strictly increasing. Since every tail atom satisfies $x_i \leq y$,

$$\sum_{i\geq 2} g(x_i) = \sum_{i\geq 2} x_i \rho(x_i) \leq \rho(y) \sum_{i\geq 2} x_i \leq (1-x)\rho(y). \tag{107}$$

Consequently

$$\bar{h} \geq L(x, y) := \frac{3}{5}x + y - \phi(x) - (1-x)\rho(y). \tag{108}$$

We next check all possible branches of L.

If $0 < y \leq x \leq 1/4$, then

$$L(x, y) = \frac{(x-y)(3-8x)}{5} \geq 0, \tag{109}$$

with equality exactly when $x = y$.

If $1/4 \leq x \leq 1/2$ and $0 < y \leq 1/4$, then

$$10L(x, y) = 1 - 2x + (16x - 6)y. \tag{110}$$

For $x \le 3/8$, the right side is minimized at $y = 1/4$, where it equals $2x - 1/2$. For $x \ge 3/8$, its infimum over $y > 0$ is the $y \downarrow 0$ value $1 - 2x$; when $x = 1/2$, the positive y-term makes the inequality strict. Hence the only zero in this case is $x = y = 1/4$.

If $1/4 \le y \le x \le 1/2$, then

$$10yL(x, y) = 10y^2 + (6x - 7)y + 1 - x. \tag{111}$$

Its derivative with respect to x is $6y - 1 > 0$. At $x = y$, the right side is $(4y - 1)^2$, so again the only zero is $x = y = 1/4$.

It remains to show strict positivity when $x \ge 1/2$. If $1/2 \le x \le 3/4$ and $0 < y \le 1/4$, then

$$10L(x, y) = 2x - 1 + (16x - 6)y > 0. \tag{112}$$

If $1/2 \le x \le 3/4$ and $1/4 \le y \le 1 - x$, put $s = 1 - x$ and $r = y/s$. Then $1/2 \le r \le 1$, $s \le 1/2$, and

$$\begin{aligned} L(x, y) &= \frac{1}{10} - s(1 - r) + \frac{1}{10r} \\ &\ge -\frac{2}{5} + \frac{r}{2} + \frac{1}{10r} \ge \frac{1}{20}. \end{aligned} \tag{113}$$

Finally, if $3/4 \le x \le 1$, then $y \le 1 - x \le 1/4$ and

$$L(x, y) = \frac{4}{5} - \frac{11}{5}x + \frac{8}{5}x^2 + \left(\frac{8}{5}x - \frac{3}{5}\right) y > 0. \tag{114}$$

Indeed, the first three terms are at least $1/20$ on this interval, and the coefficient of y is positive.

Equations (109)–(114) imply that a zero with $y > 0$ must satisfy $x = y \le 1/4$. Equality is also required throughout (107). Strict monotonicity of ρ forces every positive tail atom to equal $y = x$, and the second inequality forces $d = 0$. Thus all positive atoms equal x, and normalization requires $x = 1/j$ for an integer $j \ge 4$. Conversely, each uniform row $((1/j)^{[j]}, 0)$ satisfies $\bar{h} = 0$.

The lower contact set when $y = 0$. There is at most one positive macroscopic atom, and

$$\bar{h} = \frac{3}{5}x - \phi(x). \tag{115}$$

Solving this equation on the four branches of (89) gives only $x = 0$ and $x = 1/2$. These are, respectively, $(\varnothing, 1)$ and $((1/2), 1/2)$. Both are zeros by direct substitution. Together with the $y > 0$ case, this proves

$$\{\bar{q} : \bar{h}(\bar{q}) = 0\} = \{(\varnothing, 1), ((1/2), 1/2), ((1/j)^{[j]}, 0) : j \ge 4\}. \tag{116}$$

O.3 Positive-inventory passage and the unique count ray

For each finite pair, literal inventory equality implies equality of the positive-atom counting measures

$$\mu_n^{\mathsf{E}} := \sum_{r=1}^{m} \sum_{i : q_{n,r,i}^{\mathsf{E}} > 0} \delta_{q_{n,r,i}^{\mathsf{E}}}, \qquad \mu_n^{\mathsf{H}} = \mu_n^{\mathsf{L}}. \tag{117}$$

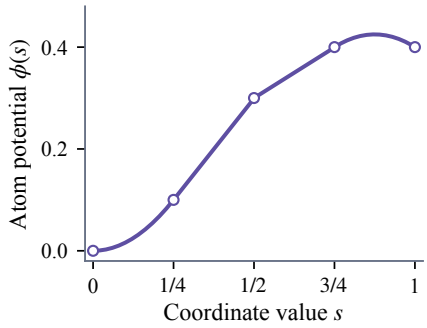
Let $f \in C_c((0, 1])$. Choose $\varepsilon > 0$ such that $\operatorname{supp} f \subseteq [\varepsilon, 1]$, and extend f by zero at the origin. A unit-mass row has at most $\lfloor 1/\varepsilon \rfloor$ coordinates at least ε. After sorting, therefore,

$$\sum_i f(q_{n,r,i}^{\mathsf{E}}) = \sum_{i=1}^{\lfloor 1/\varepsilon \rfloor} f(q_{n,r,i}^{\mathsf{E}}). \tag{118}$$

The right side is a fixed finite sum. Coordinatewise convergence and (117) now give

$$\sum_{r,i} f(x_{r,i}^{\mathsf{H}}) = \sum_{r,i} f(x_{r,i}^{\mathsf{L}}). \tag{119}$$

(a)



(b)

$\bar{h}$	Spectrum	Range
1/5	$(s, 1-s)$	$1/2 \le s \le 1$
	(s, s) + dust	$1/4 \le s \le 1/2$
0	All dust	
	(1/2) + half dust	
	$(1/j)^{[j]}$	$j \ge 4$

Figure 12: **Potential and residual contact geometry.** (a) The exact separable atom potential in Equation (89) has knots at $1/4, 1/2, 3/4$. (b) The extended residual $\bar{h}$ on dusty spectra equals $1/5$ on the two upper contact families and 0 on the three lower families shown. Matching positive inventories then reduces an extremizing limit to the five endpoint row types $U, V; W, X, Z$. The plot visualizes only the one-coordinate potential, not the dimension of the contact sets.

Table 20: Positive-atom inventory reduces the residual contact sets to five types. Their multiplicities and maxima force the primitive count ray in Equation (121).

side	type	macroscopic atoms	dust	F_1	mult. of $1/2$	mult. of $1/4$
H	U	$(1/2, 1/2)$	0	$1/2$	2	0
H	V	$(1/4, 1/4)$	$1/2$	$1/4$	0	2
L	W	$(1/2)$	$1/2$	$1/2$	1	0
L	X	$(1/4)^{[4]}$	0	$1/4$	0	4
L	Z	$\varnothing$	1	0	0	0

Thus the two limiting positive-atom counting measures are equal as locally finite Radon measures on $(0, 1]$. In particular, they have the same multiplicity at every positive atom value. Dust is invisible to these measures, exactly as intended.

By (116), every positive atom on the low side belongs to

$$S := \{1/2\} \cup \{1/j : j \ge 4\}. \tag{120}$$

The positivity of the counting measures prevents cancellation. Hence every high atom also belongs to S. A binary member $((s, 1-s), 0)$ of the upper contact set has $s \ge 1/2$; if $s > 1/2$, its s-atom is not in S, so only $s = 1/2$ is possible. An equal-head member $((s, s), 1-2s)$ must satisfy

$$s \in [1/4, 1/2] \cap S = \{1/4, 1/2\}.$$

Therefore the high rows are exactly of types U, V in Table 20. Their aggregate positive support is contained in $\{1/2, 1/4\}$, so (119) excludes every low uniform type except $j = 4$. The low rows are exactly W, X, Z.

Let n_U, n_V, n_W, n_X, n_Z count these types. The two row totals and equality of the counting measures at $1/2$ and $1/4$ give

$$n_U + n_V = m, \quad n_W + n_X + n_Z = m, \quad 2n_U = n_W, \quad 2n_V = 4n_X.$$

Equation (99) becomes

$$\frac{n_U}{2} + \frac{n_V}{4} = \frac{n_W}{2} + \frac{n_X}{4}.$$

Substituting $n_V = m - n_U$, $n_W = 2n_U$, and $n_X = n_V/2$ gives

$$\frac{m}{4} + \frac{n_U}{4} = \frac{m}{8} + \frac{7n_U}{8},$$

so $n_U = m/5$. It follows successively that

$$n_V = \frac{4m}{5}, \qquad n_W = \frac{2m}{5}, \qquad n_X = \frac{2m}{5}, \qquad n_Z = \frac{m}{5}. \tag{121}$$

Equivalently, $(n_U, n_V; n_W, n_X, n_Z) = \frac{m}{5}(1, 4; 2, 2, 1)$ for every rowwise dusty accumulation point.

Suppose $5 \nmid m$ and $\Gamma^{\mathsf{sp}}_m = 1/5$. By the definition of the supremum, an admissible sequence has absolute gaps tending to $1/5$. Swapping its sides where necessary orients the gap. Dusty compactness gives a convergent subsequence, but (121) would force the integer n_U to equal $m/5$, a contradiction. Hence

$$5 \nmid m \quad \Longrightarrow \quad \Gamma^{\mathsf{sp}}_m < \frac{1}{5}. \tag{122}$$

This is a strict fixed-m conclusion; no value of the left side is asserted here.

O.4 Sufficiency for multiples of five

It remains to exhibit a spectral sequence for $m = 5\ell$. For an integer $j \geq 2$, set

$$t = \frac{1}{2j}, \qquad a = \frac{1}{2} - t, \qquad b = \frac{1}{4}, \tag{123}$$

and let $c^{[r]}$ denote r repetitions of the entry c. Define five high rows by

$$\begin{aligned} H_1 &= (a, a, t, t), \\ H_2 = H_3 = H_4 = H_5 &= (b, b, t^{[j]}), \end{aligned} \tag{124}$$

and five low rows by

$$\begin{aligned} L_1 = L_2 &= (a, t^{[j+1]}), \\ L_3 = L_4 &= (b, b, b, b), \\ L_5 &= (t^{[2j]}). \end{aligned} \tag{125}$$

The normalization identities are

$$2a + 2t = 1, \qquad 2b + jt = 1, \qquad a + (j+1)t = 1, \qquad 4b = 1, \qquad 2jt = 1. \tag{126}$$

Both sides have the same positive occurrence inventory,

$$\{a^{[2]},\, b^{[8]},\, t^{[4j+2]}\}. \tag{127}$$

Indeed, the high t-multiplicity is $2 + 4j$, while the low t-multiplicity is $2(j+1) + 2j$. Padding every row to dimension $2j$ also makes the zero multiplicities equal.

For $j \geq 2$, the displayed leading entries are maxima, with harmless ties at $j = 2$. The two unnormalized maximum sums are

$$\sum_{r=1}^{5} F_1(H_r) = a + 4b = \frac{3}{2} - t, \tag{128}$$

and

$$\sum_{r=1}^{5} F_1(L_r) = 2a + 2b + t = \frac{3}{2} - t. \tag{129}$$

The corresponding top-two sums are

$$\sum_{r=1}^{5} F_2(H_r) = 2a + 8b = 3 - 2t, \tag{130}$$

and

$$\sum_{r=1}^{5} F_2(L_r) = 2(a + t) + 4b + 2t = 2 + 2t. \tag{131}$$

Therefore

$$\frac{1}{5}\sum_{r=1}^{5}\big(F_2(H_r)-F_2(L_r)\big)=\frac{1}{5}-\frac{2}{5j}\longrightarrow\frac{1}{5}. \tag{132}$$

Now let $m=5\ell$ and repeat every row in (124)–(125) exactly ℓ times. Both occurrence inventories and both unnormalized maximum sums are multiplied by ℓ, while division by 5ℓ leaves the average R_2 gap in (132) unchanged. Thus

$$\Gamma^{\mathsf{sp}}_{5\ell}\geq\frac{1}{5}.$$

The residual range $0\leq h\leq 1/5$ gives $\Gamma^{\mathsf{sp}}_{m}\leq 1/5$ for every m, so

$$\Gamma^{\mathsf{sp}}_{5\ell}=\frac{1}{5}. \tag{133}$$

Applying the cell-count common-prior bridge to each finite member of this sequence produces two channels with one common prior and exactly 5ℓ outputs of mass $1/(5\ell)$, preserving the posterior spectra and the gap. Together with (87), this proves $\Gamma^{\mathsf{ch}}_{5\ell}=1/5$ and completes the proof of Theorem N.2.

P Full Proof of the Exact-m Common-Prior Realization

After applying common zero padding if needed, let $A=(a_{ri})$ and $B=(b_{sj})$ be two $m\times N$ nonnegative arrays whose rows sum to one and whose pooled coordinate occurrences agree:

$$\{\{a_{ri}:r\in[m],i\in[N]\}\}=\{\{b_{sj}:s\in[m],j\in[N]\}\}. \tag{134}$$

We prove that within-row permutations produce arrays $\widehat{A},\widehat{B}$ satisfying

$$\sum_{r=1}^{m}\widehat{a}_{r\ell}=\sum_{s=1}^{m}\widehat{b}_{s\ell}\qquad(\ell\in[N]), \tag{135}$$

and hence the common prior

$$p_\ell=\frac{1}{m}\sum_{r}\widehat{a}_{r\ell}=\frac{1}{m}\sum_{s}\widehat{b}_{s\ell}. \tag{136}$$

Full proof of Lemma N.1. **Step 1: occurrence bijection.** Literal multiset equality (134) gives a value-preserving bijection

$$\beta:[m]\times[N]\longrightarrow[m]\times[N], \tag{137}$$

$$\beta(r,i)=(s,j)\quad\Longrightarrow\quad a_{ri}=b_{sj}. \tag{138}$$

For each numerical value, choose any bijection between its occurrences in A and B; their disjoint union gives β.

Step 2: bipartite multigraph. Create left vertices $A_1,\dots,A_m$ and right vertices $B_1,\dots,B_m$. For each $\beta(r,i)=(s,j)$, add a distinct edge from A_r to B_s with weight

$$w(e):=a_{ri}=b_{sj}. \tag{139}$$

Each vertex has degree N, so G is an N-regular bipartite multigraph.

Step 3: decompose into N perfect matchings. For any left set S, all $N|S|$ incident edges terminate in the neighborhood $\Gamma(S)$, giving

$$N|S|\leq N|\Gamma(S)|. \tag{140}$$

Hall's condition holds, so a perfect matching exists. Removing it leaves an $(N-1)$-regular graph; induction yields

$$E(G)=M_1\,\dot\cup\cdots\dot\cup\,M_N. \tag{141}$$

Step 4: matching colors as coordinate labels. Color every edge in M_ℓ with ℓ. Each row vertex receives each color exactly once. For the unique color-ℓ edge at A_r, set $\widehat{a}_{r\ell}$ to its A-side value; define $\widehat{b}_{s\ell}$ analogously. This permutes coordinates within every row, and edgewise value equality gives

$$\sum_{r=1}^{m} \widehat{a}_{r\ell} = \sum_{e \in M_\ell} w(e) = \sum_{s=1}^{m} \widehat{b}_{s\ell}. \tag{142}$$

Step 5: common prior and channels. Define p by (136). Row normalization gives

$$\sum_{\ell=1}^{N} p_\ell = \frac{1}{m} \sum_{r=1}^{m} \sum_{\ell=1}^{N} \widehat{a}_{r\ell} = 1. \tag{143}$$

If $p_\ell = 0$, then $\widehat{a}_{r\ell} = \widehat{b}_{s\ell} = 0$ for all r, s. Define channels

$$W^{\mathsf{A}}(r \mid \ell) := \begin{cases} \dfrac{\widehat{a}_{r\ell}}{m p_\ell}, & p_\ell > 0, \\ \dfrac{1}{m}, & p_\ell = 0, \end{cases} \qquad W^{\mathsf{B}}(s \mid \ell) := \begin{cases} \dfrac{\widehat{b}_{s\ell}}{m p_\ell}, & p_\ell > 0, \\ \dfrac{1}{m}, & p_\ell = 0. \end{cases} \tag{144}$$

Both are stochastic kernels. Under prior p, every output has marginal

$$\Pr_{\mathsf{A}}(Y = r) = \sum_{\ell=1}^{N} p_\ell W^{\mathsf{A}}(r \mid \ell) = \frac{1}{m} \sum_{\ell : p_\ell > 0} \widehat{a}_{r\ell} = \frac{1}{m}, \tag{145}$$

using row normalization and vanishing zero-prior entries. Bayes' rule with denominator $1/m$ gives

$$\Pr_{\mathsf{A}}(X = \ell \mid Y = r) = \widehat{a}_{r\ell} \tag{146}$$

for every ℓ (the zero-prior case holds because both sides are zero). The posterior vector at output r is exactly $\widehat{a}_{r\cdot}$, and likewise $\widehat{b}_{s\cdot}$ for the second channel.

If zero-probability latent states are excluded, restrict to $I_+ = \{\ell : p_\ell > 0\}$. The removed coordinates are common all-zero columns; every permutation- and padding-invariant functional is preserved.

Step 6: reverse inclusion. Conversely, two channels counted by Γ_m^{ch} share one finite latent dimension with equal indexed profiles, so lemma M.1 gives equal posterior intensities. Because every output has mass $1/m$, multiplying by m gives literal pooled occurrence equality. Their arrays are feasible for Γ_m^{sp} with the same R_1, R_2. The forward construction completes the identity $\Gamma_m^{\mathsf{ch}} = \Gamma_m^{\mathsf{sp}}$. □

Remark P.1 (Exact scope of the bridge). The bridge uses occurrence-inventory equality, not moments alone. It applies to the uniform empirical law with exactly m outputs per side; it does not turn an arbitrary Borel posterior law into an exact-m channel. No equality of R_1 is needed for the realization itself. The common prior is pair-dependent and may be nonuniform.

Corollary P.2 (Operational representation envelope)**.** *Let $Z_{\mathsf{A}}, Z_{\mathsf{B}}$ be representations of the same finite-label variable X, and write $Q_{\mathsf{E}} = \Pr\{X = \cdot \mid Z_{\mathsf{E}}\}$. If*

$$\mathrm{Law}(Q_{\mathsf{A},i}) = \mathrm{Law}(Q_{\mathsf{B},i}) \quad \textit{for every labeled class } i, \qquad \mathrm{Acc}_1^*(X \mid Z_{\mathsf{A}}) = \mathrm{Acc}_1^*(X \mid Z_{\mathsf{B}}), \tag{147}$$

then

$$|\mathrm{Acc}_2^*(X \mid Z_{\mathsf{A}}) - \mathrm{Acc}_2^*(X \mid Z_{\mathsf{B}})| \le \frac{1}{5}. \tag{148}$$

For every $\varepsilon > 0$, finite-output pairs under one common uniform class prior match every labeled coordinate marginal and Bayes top-1 accuracy, yet have a top-2 gap greater than $1/5 - \varepsilon$. The same upper bound and global sharp constant hold for the weaker requirement of equal complete Petz/down-arrow profiles and equal top-1 accuracy.

Proof. Coordinate-marginal equality implies pooled-intensity equality by Equation (11), so the first upper bound is theorem B.1. The profile upper bound is theorem Q.1 and lemma M.1; both sharpness statements follow from theorem B.4 and its common-uniform-prior realization. □

Q Robustness to Approximate and Finite-Order Summaries

To compare approximate reports across dimensions, express the potential through the probability measure λ_μ. For $x > 0$, define $\psi(x) = \phi(x)/x$, and set $\psi(0) = 0$. From Equation (26),

$$\psi(x) = \begin{cases} \frac{8}{5}x, & 0 \le x \le \frac{1}{4}, \\ \frac{4}{5} - \frac{1}{10x}, & \frac{1}{4} \le x \le \frac{1}{2}, \\ \frac{2}{5} + \frac{1}{10x}, & \frac{1}{2} \le x \le \frac{3}{4}, \\ \frac{14}{5} - \frac{8}{5}x - \frac{4}{5x}, & \frac{3}{4} \le x \le 1. \end{cases} \tag{149}$$

The pieces agree at the knots. Direct inspection gives

$$\|\psi\|_\infty = \frac{3}{5}, \qquad \mathrm{Lip}(\psi) = \frac{8}{5}. \tag{150}$$

Indeed, the maximum value $3/5$ occurs at $x = 1/2$, while the four interior derivatives are

$$\frac{8}{5}, \qquad \frac{1}{10x^2}, \qquad -\frac{1}{10x^2}, \qquad -\frac{8}{5} + \frac{4}{5x^2};$$

their absolute values are at most $8/5$, with equality on the first branch. Since $\phi(0) = 0$ and $\lambda_\mu = x\nu_\mu$,

$$R_2(\mu) = \frac{2}{5}R_1(\mu) + \int \psi \, \mathrm{d}\lambda_\mu + \mathbb{E}_\mu h. \tag{151}$$

For probability measures on $[0, 1]$, let W_1 denote one-Wasserstein distance for the cost $|x - y|$.

Theorem Q.1 (Continuity of the ambiguity envelope)**.** *Let $\mu_{\mathsf{A}} \in \mathcal{P}(\Delta_{N_{\mathsf{A}}})$ and $\mu_{\mathsf{B}} \in \mathcal{P}(\Delta_{N_{\mathsf{B}}})$, where the finite dimensions may differ. Then*

$$\boxed{\begin{aligned} |R_2(\mu_{\mathsf{A}}) - R_2(\mu_{\mathsf{B}})| \le \frac{1}{5} + \frac{2}{5}|R_1(\mu_{\mathsf{A}}) - R_1(\mu_{\mathsf{B}})| \\ + \frac{8}{5}W_1(\lambda_{\mu_{\mathsf{A}}}, \lambda_{\mu_{\mathsf{B}}}). \end{aligned}} \tag{152}$$

Consequently, equal complete profiles and equal R_1 imply an R_2 gap at most $1/5$ without a known common dimension.

Proof. Apply Equation (151) to both laws. The residual expectations differ by at most $1/5$, and $\mathrm{Lip}(\psi) = 8/5$ bounds the integral difference by $(8/5)W_1$. The exact-profile consequence uses lemma M.1. □

The transport term compares the full distributions of true-class confidence. A finite panel of profile values need not determine this distribution, but its entropy errors bound finitely many moments. The next result converts those moment bounds into a containment interval.

Corollary Q.2 (Finite-order containment)**.** *For an integer $L \ge 1$, let*

$$\delta_L := \max_{2 \le n \le L+1} |H_n^{\downarrow}(\mu_{\mathsf{A}}) - H_n^{\downarrow}(\mu_{\mathsf{B}})|.$$

Then

$$\boxed{\begin{aligned} |R_2(\mu_{\mathsf{A}}) - R_2(\mu_{\mathsf{B}})| \le \frac{1}{5} + \frac{2}{5}|R_1(\mu_{\mathsf{A}}) - R_1(\mu_{\mathsf{B}})| + \frac{8}{5\sqrt{L}} \\ + \frac{3}{5}L3^L\delta_L. \end{aligned}} \tag{153}$$

Thus exact matching through order $L + 1$, together with equal R_1, costs only $8/(5\sqrt{L})$ beyond the irreducible $1/5$. If all integer-order errors are at most $0 < \delta \le 1/9$, choosing $L = \lfloor \frac{1}{2}\log_3(1/\delta) \rfloor$ gives a vanishing additional modulus as $\delta \downarrow 0$. The coefficient multiplying δ_L is a conservative moment-inversion condition number, not an optimal statistical rate. This is a population containment, not an estimation rate, and no fixed-panel lower theorem is claimed.

Proof of corollary Q.2. Write $\lambda_s = \lambda_{\mu_s}$ for $s \in \{\mathsf{A}, \mathsf{B}\}$ and $\sigma = \lambda_{\mathsf{A}} - \lambda_{\mathsf{B}}$. For $1 \le r \le L$,

$$\int x^r \, \lambda_s(\,\mathrm{d}x) = M_{r+1}(\mu_s), \qquad \log M_{r+1}(\mu_s) = -r H^{\downarrow}_{r+1}(\mu_s) \le 0.$$

Because the exponential has derivative at most one on $(-\infty, 0]$,

$$\left| \int x^r \, \sigma(\,\mathrm{d}x) \right| = |M_{r+1}(\mu_{\mathsf{A}}) - M_{r+1}(\mu_{\mathsf{B}})| \le r\delta_L \le L\delta_L. \tag{154}$$

Let

$$(B_L\psi)(x) = \sum_{k=0}^{L} \psi\left(\frac{k}{L}\right) \binom{L}{k} x^k (1-x)^{L-k}$$

be the degree-L Bernstein polynomial. If $K_x \sim \mathrm{Bin}(L, x)$, then

$$|(B_L\psi)(x) - \psi(x)| \le \frac{8}{5}\mathbb{E}\left|\frac{K_x}{L} - x\right| \le \frac{8}{5}\sqrt{\frac{x(1-x)}{L}} \le \frac{4}{5\sqrt{L}}.$$

Since both λ_{A} and λ_{B} are probability measures,

$$\left| \int \psi \, \mathrm{d}\sigma \right| \le \frac{8}{5\sqrt{L}} + \left| \int B_L\psi \, \mathrm{d}\sigma \right|. \tag{155}$$

The $k = 0$ term vanishes because $\psi(0) = 0$. Expanding $x^k(1-x)^{L-k}$, using $\|\psi\|_\infty = 3/5$, and applying Equation (154) gives

$$\begin{aligned} \left| \int B_L\psi \, \mathrm{d}\sigma \right| &\le \frac{3}{5} L\delta_L \sum_{k=1}^{L} \binom{L}{k} \sum_{j=0}^{L-k} \binom{L-k}{j} \\ &= \frac{3}{5} L(3^L - 2^L)\delta_L \le \frac{3}{5} L 3^L \delta_L. \end{aligned} \tag{156}$$

Combining Equations (151), (155) and (156) with the residual range proves Equation (153). If a uniform all-order error is $0 < \delta \le 1/9$ and $L = \lfloor \frac{1}{2}\log_3(1/\delta) \rfloor$, then $L \ge 1$, $3^L\delta \le \sqrt{\delta}$, and both additional terms vanish as $\delta \downarrow 0$. □